

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4 p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 2, 2003, 15:46:39 ; Search time 62.6849 seconds  
(without alignments)  
1768.599 Million cell updates/sec

Title: US-09-079-678-178

Perfect score: 4321

Sequence: 1 MTLQHLHLSCLLMLYLATG.....DRGKDNVSAQAQSEVKPLRS 832

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_101002.\*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4321	100.0	832	20	Human HPT-1 protei
2	4321	100.0	832	20	C880P similar amin
3	908.5	21.0	807	21	Human PRO1340 (UNQ
4	908.5	21.0	807	22	Human PRO1340. Ho
5	908.5	21.0	807	22	Protein of the inv
6	908.5	21.0	807	23	Human PRO protein,
7	877.5	20.3	830	22	Human cadherin-2 (
8	699.5	16.2	906	21	Human N-cadherin.
9	699.5	16.2	906	22	Novel human diagno
10	699.5	16.2	906	23	Human N-cadherin p

11	696	16.1	912	23	ABB81471	Chicken N-cadherin
12	694.5	16.1	906	23	ABB57233	Mouse ischaemic co
13	684.5	15.8	916	18	AAW25658	Human cadherin-4.
14	684.5	15.8	916	18	AAW13129	Full length human
15	671.5	15.5	896	15	AAW63533	Human Ht-1376 cell
16	671.5	15.5	896	23	AAU78055	Human desmocollin
17	667.5	15.4	847	22	AAW39235	Human polypeptide
18	665.5	15.4	862	22	AAW41021	Human polypeptide
19	657.5	15.2	713	18	AAW25638	Human cadherin-13.
20	657.5	15.2	713	18	AAW13136	Full length human
21	630.5	14.6	712	13	AAW27823	Sequence encoded b
22	630.5	14.6	717	13	AAW27824	Rat fat 3 protein
23	611	14.1	4555	23	AAW52106	Novel human diagno
24	607	14.0	759	22	ABG21947	Novel human diagno
25	605	14.0	4591	22	ABG22977	Mouse E-cadherin p
26	604	14.0	889	23	ABB81472	Mouse E-cadherin p
27	601.5	13.9	3606	22	ABW62595	Drosophila melanog
28	593	13.7	878	15	AAW55060	Sequence of human
29	593	13.7	878	16	AAW85487	Human E-cadherin p
30	593	13.7	878	23	ABB81475	Human E-cadherin p
31	592.5	13.7	117	21	AAV64597	Nonclassical cadhe
32	591.5	13.7	4643	22	ABW71609	Drosophila melanog
33	591	13.7	882	20	AAW09375	Wild-type E-cadher
34	591	13.7	882	21	AAW35730	Human E-cadherin.
35	591	13.7	882	22	AAW73490	Human E-cadherin p
36	591	13.7	882	23	AAU78051	Human novel cytoki
37	576	13.3	3014	22	AAU68533	Seven-pass transme
38	576	13.3	3014	22	AAU02196	Human NOV7 protein
39	576	13.3	3028	22	AAW08586	Mouse PRP-binding
40	574.5	13.3	796	22	AAW60418	Sequence encoded b
41	573.5	13.3	796	15	AAW49730	Human protein SEQ
42	571.5	13.2	788	22	AAW78375	Mouse ischaemic co
43	571.5	13.2	796	23	ABB57294	Mouse ischaemic co
44	570.5	13.2	840	22	ABG01693	Novel human diagno
45	569.5	13.2	840	22	AAW23903	Human EST encoded

## ALIGNMENTS

RESULT 1

AAW74089

ID AAW74089 standard; Protein; 832 AA.

XX AC AAW74089;

XX AC AAW74089;

DT 04-MAY-1999 (first entry)

XX Human HPT-1 protein sequence.

XX Gastro-intestinal transport receptor; binding protein; hSI; HPT1;

XX D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;

XX Intestinal peptide-associated transporter; hypertension; diabetes;

XX osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;

XX therapeutic agent delivery; therapy; probe.

XX Homo sapiens.

XX WO9851325-A2.

XX 19-NOV-1998.

XX 15-MAY-1998; 98WO-US10088.

XX 15-MAY-1997; 97US-0046595.

XX (CYTO-) CYTOGEN CORP.

XX (ELAN-) ELAN CORP PLC.

XX Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin LJ;

XX Onahony DJ, Patterson CA, Singleton J;

XX WPI; 1999-009568/01.

DR

DR N-PSDB; AAX18166.

XX New proteins that bind specifically to receptors in the

PT gastro-intestinal tract and related nucleic acid - chimaeras and

PT antibodies, used to deliver therapeutic or diagnostic agents to, or

PT through, the gastrointestinal tract, e.g. insulin or leuprolide

XX

PS Disclosure; Fig 2; 294pp; English.

XX

CC This sequence is the human hPT-1 protein. The invention relates to

CC purified proteins (I) that bind specifically to at least one of the

CC gastro-intestinal (GI) tract receptors human intestinal

CC peptide-associated transporter (hPT1), hPEPT1, D2H and human

CC sucrose-isomaltase complex (hSI). (I) provide active transport of

CC therapeutic agents through human and animal GI tissue (into the blood)

CC for in vivo delivery, particularly for treatment or prevention

CC of hypertension, diabetes, osteoporosis, haemophilia, anaemia, cancer,

CC migraine, or angina pectoris. Specifically they are used to deliver

CC insulin or leuprolide, but many other suitable therapeutic agents are

CC disclosed, including genes or inhibitory nucleic acid, imaging agents and

CC antigens. (I) may also provide targeting to the GI tract. Other uses of

CC (I) are: (i) to determine the level of specified receptors in a sample

CC (in a binding assay); and (ii) to screen for molecules that bind (I).

CC Immunogenic analogues or derivatives of (I) are used to raise antibodies

CC and in immunoassays. The antibodies are used to locate, detect and

CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis

CC etc., also for peptide purification and immobilisation.

XX

Sequence 832 AA;

Query Match 100.0%; Score 4321; DB 20; Length 832;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 832; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTLQAHLSLCLMLYLATGYGQEGKFSGLPKPMTFSYIEGQEPSQIIFQFKANPPAVTF 60

Db 1 MTLQAHLSLCLMLYLATGYGQEGKFSGLPKPMTFSYIEGQEPSQIIFQFKANPPAVTF 60

QY 61 ELTGTDNIFVIEREGLLYYNLRALDRETRSTHNLQVAALDANGIIVEGVPITIEVKDIN 120

Db 61 ELTGTDNIFVIEREGLLYYNLRALDRETRSTHNLQVAALDANGIIVEGVPITIEVKDIN 120

QY 121 DNRPTFLASKYEGSVQRNRPKPELYNATDLPDPATNGOLYQYIVQLPMMNNMYF 180

Db 121 DNRPTFLASKYEGSVQRNRPKPELYNATDLPDPATNGOLYQYIVQLPMMNNMYF 180

QY 181 QINNTGTGALSITREGSQELNPAKNPSYNLVTSVKDMGQSGNSFSDTTSVDIIVTENIWK 240

Db 181 QINNTGTGALSITREGSQELNPAKNPSYNLVTSVKDMGQSGNSFSDTTSVDIIVTENIWK 240

QY 241 APKPVEMVENSTDHPPIKITQVRNMDPCAQYSLVDKEKLPRFPFSDIDQEGDIYVTPQLDR 300

Db 241 APKPVEMVENSTDHPPIKITQVRNMDPCAQYSLVDKEKLPRFPFSDIDQEGDIYVTPQLDR 300

QY 301 BEKDAYFYAVAKDYGRKPLSYPLBIHVKKVDINDNPTCPSPVTVFEQENRGLNSIG 360

Db 301 BEKDAYFYAVAKDYGRKPLSYPLBIHVKKVDINDNPTCPSPVTVFEQENRGLNSIG 360

QY 361 TLTAHREDEENTANSFLNRYVEQTPKPLMDGLFLIQTAYAGMLQAKSLKKQDTPQYNL 420

Db 361 TLTAHREDEENTANSFLNRYVEQTPKPLMDGLFLIQTAYAGMLQAKSLKKQDTPQYNL 420

QY 421 TIEVSDKDFKTLCFQYINVIDINDQIPFEKSDYGNLTAEEDTNGISTILITQATDADEP 480

Db 421 TIEVSDKDFKTLCFQYINVIDINDQIPFEKSDYGNLTAEEDTNGISTILITQATDADEP 480

QY 481 FTGSKILYHIKDGSEGRGLVDTPHTNTGVYIHKPLDPEETAASVNIYFKAENPEPLV 540

Db 481 FTGSKILYHIKDGSEGRGLVDTPHTNTGVYIHKPLDPEETAASVNIYFKAENPEPLV 540

QY 541 FGKYNASSFAKFTLIVTDVNEAPOSQHVQAKVSEDAVCTKGVNTAKDPEGLDISY 600

Db 541 FGKYNASSFAKFTLIVTDVNEAPOSQHVQAKVSEDAVCTKGVNTAKDPEGLDISY 600

QY 601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRQVQVATEYGGSSLSVSEFHLIMDV 660

Db 601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRQVQVATEYGGSSLSVSEFHLIMDV 660

QY 661 NDNPPRLAKDYTGTLFCHPLSAPGSLIFPATDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720

Db 661 NDNPPRLAKDYTGTLFCHPLSAPGSLIFPATDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720

QY 721 INGTCHARLSTRHTDEERAYVVLIRINDGGRPLEGIVSLPVTFCSCVSGSCFRPAGHOT 780

Db 721 INGTCHARLSTRHTDEERAYVVLIRINDGGRPLEGIVSLPVTFCSCVSGSCFRPAGHOT 780

QY 781 GIPTVGMAGVILLTLLVIGIILAVVFIKDKGKDNVESQAQASVKPLRS 832

Db 781 GIPTVGMAGVILLTLLVIGIILAVVFIKDKGKDNVESQAQASVKPLRS 832

RESULT 2

AAM24522

ID AAM24522 standard; Protein; 832 AA.

XX AAM24522;

AC AAM24522;

XX 12-OCT-2001 (first entry)

DT C880P similar amino acid sequence (GENESEQ W740898).

XX Human; immunotherapy; diagnosis; colon cancer; colon tumour;

KW immunogenic; gene therapy; vaccine; colon cancer.

OS Homo sapiens.

XX WO2000149716-A2.

XX 12-JUL-2001.

PF 29-DEC-2000; 2000WO-US35596.

XX 30-DEC-1999; 99US-0476296.

PR 10-JAN-2000; 2000US-0480321.

PR 15-FEB-2000; 2000US-0504629.

PR 06-MAR-2000; 2000US-0519444.

PR 19-MAY-2000; 2000US-0575251.

PR 29-JUN-2000; 2000US-0609448.

PR 28-AUG-2000; 2000US-0649811.

XX (CORI-) CORIXA CORP.

PA Xu J, Lodes MJ, Secret H, Benson DR, Meagher MJ, Stolk JA;

PI King GE, Wang T, Jiang Y;

XX WPI; 2001-441847/47.

XX Colon tumor associated proteins and nucleic acids useful for the

PT prevention, diagnosis and treatment of colonic cancer -

XX Claim 2; Page 469-472; 472pp; English.

The present invention describes colon tumour associated proteins (I) and the polynucleotides (II) that encode them. (I) have cytostatic activity. (I) and (II) can be used in gene therapy and vaccine production. (I) and (II) may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate colon tumour associated protein (TCAP) expression, such as colonic cancer. For example, (I) and (II) may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of TCAPs by expressing inactive proteins or to supplement the patients own production of them. Additionally, (II) may be used to produce the TCAP proteins, by inserting the nucleic acids into a host cell culturing the cell to express the protein. (II) and its complementary sequences may also be used as DNA probes in diagnostic polymerase chain reaction (PCR) and hybridisation assays to detect and





PR	29-SEP-1998;	98US-0102207.	XX	Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
PR	29-SEP-1998;	98US-0102240.	PI	
PR	29-SEP-1998;	98US-0102307.	XX	
PR	29-SEP-1998;	98US-0102330.	DR	WPI: 2000-237871/20.
PR	29-SEP-1998;	98US-0102331.	DR	N-PSDB; AAA37087.
PR	30-SEP-1998;	98US-0102484.	XX	
PR	30-SEP-1998;	98US-0102487.	PT	New mammalian DNA sequences encoding transmembrane, receptor or
PR	30-SEP-1998;	98US-0102571.	PT	secreted PRO polypeptides, useful for screening of potential peptide or
PR	01-OCT-1998;	98US-0102684.	PT	small molecule inhibitors of the relevant receptor/ligand interactions
PR	01-OCT-1998;	98US-0102687.	XX	
PR	02-OCT-1998;	98US-0102965.	PS	Claim 12; Fig 132; 773pp; English.
PR	06-OCT-1998;	98US-0103258.	XX	
PR	06-OCT-1998;	98US-0103449.	CC	AAA37022 to AAA37144 encode the new isolated human transmembrane,
PR	07-OCT-1998;	98US-0103314.	CC	receptor or secreted PRO polypeptides given in AAY99340 to AAY99462. The
PR	07-OCT-1998;	98US-0103315.	CC	transmembrane and receptor PRO proteins can be used for screening of
PR	07-OCT-1998;	98US-0103328.	CC	potential peptide or small molecule inhibitors of the relevant
PR	07-OCT-1998;	98US-0103395.	CC	receptor/ligand interactions. The polypeptides and nucleotide sequences
PR	07-OCT-1998;	98US-0103396.	CC	encoding then have various industrial applications, including uses as
PR	07-OCT-1998;	98US-0103401.	CC	pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
PR	08-OCT-1998;	98US-0103633.	CC	PCR primers and hybridisation probes used in the isolation of the PRO
PR	08-OCT-1998;	98US-0103678.	CC	polypeptides from the present invention.
PR	08-OCT-1998;	98US-0103679.	XX	
PR	08-OCT-1998;	98US-0103711.	SQ	Sequence 807 AA;
PR	14-OCT-1998;	98US-0104257.		
PR	20-OCT-1998;	98US-0104987.		Query Match 21.0%; Score 908.5; DB 21; Length 807;
PR	20-OCT-1998;	98US-0105000.		Best Local Similarity 30.2%; Pred. No. 1.1e-64;
PR	20-OCT-1998;	98US-0105002.		Matches 237; Conservative 159; Mismatches 335; Indels 53; Gaps 21;
PR	21-OCT-1998;	98US-0105104.	Qy	62 LTGE----TDNIFVIERE-GLLYNRLDRETRSTHNNQVAALDANGLIIVSGPPIITIEV 116
PR	22-OCT-1998;	98US-0105169.	Db	: : : : :       : : : : :     : : : : :     : : : : :     : : : : :
PR	22-OCT-1998;	98US-0105266.		57 LSGDSGRATEPGFAMDPDSGFLLVTRALDREEQAEYQLQVLEMDQGHVLMGPQVPLVHV 116
PR	26-OCT-1998;	98US-0105693.	Qy	: : : : :       : : : : :     : : : : :     : : : : :     : : : : :
PR	26-OCT-1998;	98US-0105694.	Db	117 KDINDNRPTFLQSKYEGSVQRNSRGKPFLLYVNATDLPATPNQQLYQIVYIQLPMINN 176
PR	27-OCT-1998;	98US-0105807.	Qy	: : : : :       : : : : :     : : : : :     : : : : :     : : : : :
PR	27-OCT-1998;	98US-0105881.	Db	117 KDENDQVPHFSQAIYRARLSRGTRGIPFLFLEASDRDEPCTANSDLRHLISQAQAPS 176
PR	27-OCT-1998;	98US-0105882.	Qy	: : : : :       : : : : :     : : : : :     : : : : :     : : : : :
PR	27-OCT-1998;	98US-0106082.		177 VMYFOINKTKAISLTREGSOELNPAKPSNLYISVKDMGQSGNSFSDTSDVDIIVTE 236
PR	28-OCT-1998;	98US-0106029.	Db	: : : : :       : : : : :     : : : : :     : : : : :     : : : : :
PR	28-OCT-1998;	98US-0106030.	Qy	177 PDMEFQLEPRLGALALSPKSTSLDHALERTYQLLVQVRMDGQA-SGHOATATVEVSIE 235
PR	28-OCT-1998;	98US-0106032.	Db	: : : : :       : : : : :     : : : : :     : : : : :     : : : : :
PR	28-OCT-1998;	98US-0106033.	Qy	237 NIWKAPKPEVMVENSTDPHPKIKITQVRWNDPGAQVSLVDKEKLPFPFSDIQEGDIYVQ 296
PR	28-OCT-1998;	98US-0106178.	Db	: : : : :       : : : : :     : : : : :     : : : : :     : : : : :
PR	29-OCT-1998;	98US-0106248.	Qy	236 STWVSLEPIHLAENLKVLYPHHMAQVHWSGGDVHYHL---ESHPPGPEVNAEGNLYTR 292
PR	29-OCT-1998;	98US-0106384.	Db	: : : : :       : : : : :     : : : : :     : : : : :     : : : : :
PR	29-OCT-1998;	98US-0108500.	Qy	297 PLDREKDAYVYAVAKDEYKPLSYPLEIHVKVDINDNPPTCPSPVTVFEVQENRGL 356
PR	30-OCT-1998;	98US-0106464.	Db	: : : : :       : : : : :     : : : : :     : : : : :     : : : : :
PR	03-NOV-1998;	98US-0106856.	Qy	293 ELDRQAQAEYLLQVRAQNSHGEDIAAPLEHLVLMENDNVNPPICPPRPTVTSIPELSPG 352
PR	03-NOV-1998;	98US-0106902.	Db	: : : : :       : : : : :     : : : : :     : : : : :     : : : : :
PR	03-NOV-1998;	98US-0106905.	Qy	357 NSIGTLTAHDRDEENTANSFLNYRIVEQTKLPMDG-LFLIQTAYAGMLQAKQSLKK-OD 414
PR	03-NOV-1998;	98US-0106919.	Db	: : : : :       : : : : :     : : : : :     : : : : :     : : : : :
PR	03-NOV-1998;	98US-0106932.	Qy	353 TEVTRLSAEDADAPGSPNSHVYQLLSPEDEGVEGRAFDPTSGSVTLGLVPLURAGN 412
PR	03-NOV-1998;	98US-0107783.	Db	: : : : :       : : : : :     : : : : :     : : : : :     : : : : :
PR	17-NOV-1998;	98US-0108775.	Qy	413 ILLVLAMDLAGAEGGFSSTCEVEAVTIDINDHAPEFITSQIGPISLPEDEVPGLVAML 472
PR	17-NOV-1998;	98US-0108779.	Db	: : : : :       : : : : :     : : : : :     : : : : :     : : : : :
PR	17-NOV-1998;	98US-0108787.	Qy	473 QATDAD-EPFTGSSKIL-YHIKGDSEGRGLGVDTPHNTGYVIK--KPLDFFFAVSN 528
PR	17-NOV-1998;	98US-0108849.	Db	: : : : :       : : : : :     : : : : :     : : : : :     : : : : :
PR	17-NOV-1998;	98US-0108801.	Qy	473 TAIDADLEP---APRLMDFAIERGDTGTEGLDWEPP--DSGHRVRLRCKNLISYEAPASHE 527
PR	17-NOV-1998;	98US-0108806.	Db	: : : : :       : : : : :     : : : : :     : : : : :     : : : : :
PR	17-NOV-1998;	98US-0108807.	Qy	529 IVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQSFQVFOAKVSEDVAICTKGVN 588
PR	17-NOV-1998;	98US-0108867.	Db	: : : : :       : : : : :     : : : : :     : : : : :     : : : : :
PR	17-NOV-1998;	98US-0108925.	Qy	528 VVVVVQSAKLIV-GPGPGGATATVTIVLVRMPPPKLDQSEYASVPSIPASAPGFLITI 586
PR	18-NOV-1998;	98US-0108848.	Db	: : : : :       : : : : :     : : : : :     : : : : :     : : : : :
PR	18-NOV-1998;	98US-0108850.	Qy	589 TAKDPEGLDISYSLRGDTRGWLKIDHVTGEIFSVAPLD-REAGSPYRVQVVAIVEGSSSL 647
PR	18-NOV-1998;	98US-0108851.	Db	: : : : :       : : : : :     : : : : :     : : : : :     : : : : :
PR	18-NOV-1998;	98US-0108852.	Qy	587 QPSDPIISRTLRFSLVNDSEGLICIEKFSGEVHTAQSLQAGPGDTYTVLVQAQDTA---- 642
PR	18-NOV-1998;	98US-0108858.	Db	: : : : :       : : : : :     : : : : :     : : : : :     : : : : :
PR	18-NOV-1998;	98US-0108904.	Qy	648 SSVSEFHILMDVNDNPRLAKDYTLGFFCHPLSAPGSLIF-EATDDQHLFRPHFTFS 706
PR	18-NOV-1998;	98US-0108904.	Db	: : : : :       : : : : :     : : : : :     : : : : :     : : : : :
PR	18-NOV-1998;	98US-0108904.	XX	643 -----LTLAPVPSQ-----YLCPTQRDHLIVSGSPSKDPLASCHGP-YSFT 683
PR			Qy	707 LG-SGSLQNDWEVSKINGTHARLSTRHTDFFERAYVVLIRINDGRRPPLEGIVSLPVTF 765

Db 684 LGPNPTVQDRWLQTLNGSHAYLTALHWVEPREHIIPVVVSHNAQ---NWQLLVRVIVC 740  
QY 766 SC-VEGSCFRPAGHQTGIPVGNAGVLTLLVIGILAVVFIKDKGKDNVESQAQ 824  
Db 741 RCNVEGCMRKVGMKGMPTKLSAVGILVGLTVAIGIFLLIFLTHWTMSRKDPDPADS 800  
QY 825 SEVK 828  
Db 801 VPLK 804

RESULT 4  
AAB87574  
ID AAB87574 standard; Protein; 807 AA.  
AC AAB87574;  
DT 15-MAY-2001 (first entry)  
DE Human PRO1340.  
KW Human; PRO protein; mapping.  
XX Homo sapiens.  
OS WO200116318-A2.  
PN 08-MAR-2001.  
XX 24-AUG-2000; 2000WO-US23328.  
XX 01-SEP-1999; 99WO-US20111.  
PR 15-SEP-1999; 99WO-US21090.  
PR 07-DEC-1999; 99US-0169495.  
PR 09-DEC-1999; 99US-0170262.  
PR 11-JAN-2000; 2000US-0175481.  
PR 18-FEB-2000; 2000WO-US04341.  
PR 18-FEB-2000; 2000WO-US04342.  
PR 22-FEB-2000; 2000WO-US04414.  
PR 01-MAR-2000; 2000WO-US05601.  
PR 03-MAR-2000; 2000US-0187202.  
PR 23-APR-2000; 2000US-0199397.  
PR 22-MAY-2000; 2000WO-US14042.  
PR 05-JUN-2000; 2000US-0209832.  
XX (GETH ) GENENTECH INC.  
PA Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;  
PI Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;  
XX WPI: 2001-183260/18.  
DR N-PSDB; AAF92106.  
XX Eighty four nucleic acids encoding PRO polypeptides, useful in  
PT molecular biology, including use as hybridization probes, and in  
PT chromosome and gene mapping. -  
XX Claim 12; Fig 98; 278pp; English.

CC The present sequence is a human PRO polypeptide (secreted and  
transmembrane). The PRO protein, and PRO agonists, PRO antagonists or  
anti-PRO antibodies are useful for preparation of a medicament useful in  
the treatment of a condition which is responsive to the PRO protein,  
agonists, antagonists or anti-PRO antibodies. The PRO protein may also be  
employed as molecular weight markers for protein electrophoresis. The PRO  
coding sequence has applications in molecular biology, including use as  
hybridisation probes, and in chromosome and gene mapping.

XX Sequence 807 AA;  
SQ

Query Match 21.0%; Score 908.5; DB 22; Length 807;  
Best Local Similarity 30.2%; Pred. No. 1.1e-64;

Matches 237; Conservative 159; Mismatches 335; Indels 53; Gaps 21;  
QY 62 LTGE---TDNIEVIERE-GLLYNRRALDRETRSTHNLQVAALDANGIIIVGGVPPIFV 116  
Db 57 LSDSGKATGEPAMDPSGFLVTRALDREEQAEYQVLTLEMDQGHVWNGQPVLVHV 116  
QY 117 KDINDNRPTFLQSKYEGSVRQNSRPGKPFYLVNATLDDPATNGQYYQIVIQPLMINN 176  
Db 117 KDENDQVPHEFSAIYRARLSRGTRPGIPFLFLEASDRDEPGTANSRLRHLFSLQAQPS 176  
QY 177 VMYFQINNKTGAISLTPREGSOELNPAKNPSYNLVISVKMGQSGENSEFSDTTSVIVTE 236  
Db 177 PDMFQLEPRIGALALSPKSTSLDHALERTYQLLVQVKMDGQA-SGHQATATVEVSII 235  
QY 237 NIWAKPKPVEMVENSTDPHPKIKITQVRWNDPGAQYSLVDKEKLPFRPFSDIOEGDIYVQ 296  
Db 236 STWVSLEPIHLAENLKVLYPHHMAQVHWSGVDVHYHL---ESHPPGPFVEAENGLVTR 292  
QY 297 PLDREKDAYVYAVAKDEYKPLSYPLEIHKVVKDINDNPPTCPSPVTVFEQENERLG 356  
Db 293 ELDRQAQAYLLQVRAQNSHGEDIAAPLELHVLMENDNVNVPICPPRDPVTSIPELSPG 352  
QY 357 NSIGTLTAHDEENTANSFLNRYVEQTPKLPMDG-LFLIQTAYAGMLQAKOSLKK-QD 414  
Db 353 TEVTRLSAEDADAPGSPNSHVYQLLSPEPEDGVGEGRAFQVDTSGSVTLGVLPDRAGN 412  
QY 415 TPQYNLTIEVSKD--FKTLCFQVINVIDINQIPIEKSQDYGNTLAEDTINIGSTILTI 472  
Db 413 ILLVLAMLAGAEGGFSSTCEVEAVTDINDHAPEITSQIGPISLPEDVEPTLVAML 472  
QY 473 QATDAD-EPFTGSKIL-YHIKDGSEGLGVDTDPHTNTGYVLIK--KPLDFFETAASVN 528  
Db 473 TAIADALEP---AFRLMDFAIERDGTGTGLDWEPI--DSGHVRLRCKNLSYEAAPSHE 527  
QY 529 IVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFQSVQVFOAKYSEDAIGTKVGNV 588  
Db 528 VVVVQSVAKLV-GPGPGCATATVTLVERVMPPPKLDQESYEASVPIISAPAGSFLTI 586  
QY 589 TAKDPEGLDISYSLRGDTRGWLKIDHVTGEIFSVAPILD-REAGSPYRVQVVATEVGGSSL 647  
Db 587 QPSDPISTRLRFSLVNDSEGNLCIEKFSGEVHTAQSLOGAQPQDGYTVLVLEAQDTA--- 642  
QY 648 SSVSEFHLILMDVNDNPPRLAKDYTGFFCHPLSAPGSLIF-EATDDQHLFRGPHFTFS 706  
Db 643 -----LTLAPVPSQ-----YLCPTPRDHLIVSGPSKDPDLASGHGP-YSFT 683  
QY 707 LG-SGSLONDWEYSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLGIVSLPVTFC 765  
Db 684 LGPNPTVQDRWLQTLNGSHAYLTALHWVEPREHIIPVVVSHNAQ---MWQLLVRVIVC 740  
QY 766 SC-VEGSCFRPAGHQTGIPVGNAGVLTLLVIGILAVVFIKDKGKDNVESQAQ 824  
Db 741 RCNVEGCMRKVGMKGMPTKLSAVGILVGLTVAIGIFLLIFLTHWTMSRKDPDPADS 800  
QY 825 SEVK 828  
Db 801 VPLK 804

RESULT 5  
AAB66154  
ID AAB66154 standard; protein; 807 AA.  
XX AAB66154;  
AC AAB66154;  
XX 02-APR-2001 (first entry)  
DT Protein of the invention #66.  
XX Secreted; transmembrane; gene therapy.  
XX Unidentified.  
OS

PN WO200078961-A1.  
 XX 28-DEC-2000.  
 XX 18-FEB-2000; 2000WO-US04342.  
 XX 23-JUN-1999; 99US-0141037.  
 PR 20-JUL-1999; 99US-0144758.  
 PR 26-JUL-1999; 99US-0145698.  
 PR 01-SEP-1999; 99WO-US20111.  
 PR 29-OCT-1999; 99US-0162506.  
 PR 30-NOV-1999; 99WO-US28313.  
 PR 02-DEC-1999; 99WO-US28551.  
 PR 16-DEC-1999; 99WO-US30095.  
 PR 05-JAN-2000; 2000WO-US00219.  
 PR 06-JAN-2000; 2000WO-US00376.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;  
 PI Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;  
 PI Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;  
 PI Watanabe CK, Williams PM, Wood WL;  
 XX  
 DR WPI; 2001-071395/08.  
 XX  
 CC Secreted and transmembrane proteins and nucleic acids designated PRO,  
 PT useful as hybridization probes, in chromosome and gene mapping and gene  
 PT therapy -  
 XX  
 XX Claim 1; Fig 132; 787pp; English.  
 XX  
 CC The present invention relates to secreted and transmembrane proteins.  
 CC These proteins and the DNA encoding them may be used as hybridization  
 CC probes, in chromosome and gene mapping and in the generation of  
 CC anti-sense RNA and DNA. They may also be used to generate either  
 CC transgenic animals or knockout animals which are in turn useful for  
 CC development and screening of therapeutically useful reagents.  
 CC The nucleic acids may also be used in gene therapy.  
 XX  
 XX Sequence 807 AA;  
 XX  
 Query Match 21.0%; Score 908.5; DB 22; Length 807;  
 Best Local Similarity 30.2%; Pred. No. 1.1e-64;  
 Matches 237; Conservative 159; Mismatches 335; Indels 53; Gaps 21;  
 QY 62 LTGE---TDNIFVIERE-GLLYNRALDRETRSTHNLQVAALDANGIIVEGVPPIPIEV 116  
 DB 57 LSGDSGKATEGPFAMDPDSGLLVTRALDREEQAEYQLVTLQMDDGHVLWGQPVLVHV 116  
 QY 117 KDINDNPTFLQSKYEGSVQRNGRPFLYVYNATDLDPATNGQLYQIVIQPLMNN 176  
 DB 117 KDENDQVPFHSQATYRARLSRGTRPGIPFLFLEASDRDEPCTANSDLRPHLSQAQPS 176  
 QY 177 VMFYQINKTGATSLTREGSOELNPAKNPSNVLVISVKMGQSGSENSFSDTSDIIVTE 236  
 DB 177 PMFQLEPRIGALALSPKSGSLDHALERTYQLLVQVQKMDGQA-SGHQATATVEVSIIE 235  
 QY 237 NIWKAPKPVENVSTDPHPKIKITQVRWNPDPGAQYSLVDKEKLPFRFPFIDQEDIYVTO 296  
 DB 236 STWVSLPEPIHLAENLKVLYPHHMAQVHWSGDDVHYHL---ESHPPGPFVNAEGNLVTR 292  
 QY 297 PLDREKDAYVFAVAKEYGKPLSYPLETHVKVKDINDNPPPCSPVTVFEVOENRGLG 356  
 DB 293 ELDRQAQAEYLLQVRAQNSHGDYAAPLEHLVLMNDNVNPPCPRPDPVTSIPELSPGP 352  
 QY 357 NSTGTLTAHRDEENTANFLNRIVEQTPKLPMDG-LFLIQTYAGMLQAKSLKK-QD 414  
 DB 353 TEVTRLSAEDADAPGSPNSHWYQLLSPEPEDGVEGRAQVQDPTSGSVTLGLVPLFRAGON 412  
 QY 415 TPQYNLTIEVSDKD--FKTLCFQVQINVIDINDQIPIFEKSDYGNLTLAEDNIGSTILTI 472  
 DB 413 ILLVLVANDLAGAGGESSSTCEVEAVTDINDHAPEITSQIGPISLPEDVEPGTLVAML 472

QY 473 QATDAD-EPETGSKIL-YHIKIGDSEGRGLGVDTPHTNTGYIHK--KPLDFETAASN 528  
 DB 473 TAJDADLEP---AFRLMDFAIERGDTGETGLDWEF--DSGHVRLRCKNLSEAPASHE 527  
 QY 529 IVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFOAKYSEDVAICTKGVN 588  
 DB 528 VVVVQSVAKLV-GPGPGGATATVTIVVERVMPPPKLDQESYEASVPIASAPAGSFLTTI 586  
 QY 589 TAKDPEGLDISYLRGDTRGWLKIDHVTGIFSVAPLD-REAGSPYRVQVVVATEVGGSSL 647  
 DB 587 QPSDPISRTLRFLSVNDSEGLCEKFSGEVHTAQSLQGAQPGDITYVLVEAQDTA---- 642  
 QY 648 SSVSEPHLIMDYNDNPPRLAKDYTLGFFCHPLSAPGSLIP-EATDDQHLFRGPHFTFS 706  
 DB 643 -----LTLAPVPSQ-----YLCPTRODHGLIVSGPSKDPDLASGHGP-YSFT 683  
 QY 707 LG-SGSLONDWEYSKINGTHARLSTRHTDFEERAYVLLIRINDGGRPPLEICVSLPTFC 765  
 DB 684 LGPNPTVQRDMRLQTLNGSHAYLTALHWPEPREHILIPVVVSHNAQ---MWQLLRVIVC 740  
 QY 766 SC-VEGSCFRPAGHQGTGPTVGMAGVILLTLLVIGILAVVFIRIKKDKGNVESQAQ 824  
 DB 741 RCNVEGOCMRKVGMRKGMPTKLSAVGILVGLTVAIGIFLILIFTHWTMSRKDPDQADS 800  
 QY 825 SEVK 828  
 DB 801 VPLK 804  
 RESULT 6  
 AAU83685  
 ID AAU83685 standard; Protein; 807 AA.  
 XX  
 AC AAU83685;  
 XX  
 DT 08-MAY-2002 (first entry)  
 XX  
 DE Human PRO protein, Seq ID No 188.  
 XX  
 KW Human; secreted protein; PRO; tumour; lung cancer; colon cancer;  
 KW breast cancer; prostate tumour; rectal tumour; liver tumour;  
 KW pericyte cell proliferation; chondrocyte cell proliferation;  
 KW tumour necrosis factor-alpha.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200208288-A2.  
 XX  
 PD 31-JAN-2002.  
 XX  
 PF 29-JUN-2001; 2001WO-US21066.  
 XX  
 PR 20-JUL-2000; 2000US-219556P.  
 PR 25-JUL-2000; 2000US-220585P.  
 PR 25-JUL-2000; 2000US-220605P.  
 PR 25-JUL-2000; 2000US-220607P.  
 PR 25-JUL-2000; 2000US-220624P.  
 PR 25-JUL-2000; 2000US-220638P.  
 PR 25-JUL-2000; 2000US-220664P.  
 PR 25-JUL-2000; 2000US-220666P.  
 PR 26-JUL-2000; 2000US-220893P.  
 PR 28-JUL-2000; 2000WO-US20710.  
 PR 23-AUG-2000; 2000WO-US23522.  
 PR 24-AUG-2000; 2000WO-US23328.  
 PR 15-SEP-2000; 2000US-000000P.  
 PR 10-NOV-2000; 2000WO-US30873.  
 PR 28-NOV-2000; 2000US-253646P.  
 PR 01-DEC-2000; 2000WO-US32678.  
 PR 20-DEC-2000; 2000US-074259.  
 PR 20-DEC-2000; 2000WO-US34956.  
 PR 28-FEB-2001; 2001WO-US06520.  
 PR 10-MAY-2001; 2001US-0854280.



PT inflammatory, immune system and cell proliferation disorders -  
XX  
PS Claim 13; Figure 8; 147pp; English.

XX This sequence represents the cadherin-2 (CDNH-2) protein of the  
CC invention. This invention comprises the nucleotide and protein  
CC sequences of two members of the cadherin family of membrane  
CC glycoproteins CDNH-1 and CDNH-2. Also included in the invention is  
CC a method for producing the CDNH proteins by culturing a transfected  
CC host cell line with a vector containing the CDNH proteins and an  
CC antibody which selectively binds to the protein. The CDNH proteins  
CC may have neurotropic, neuroprotective, cardiac, antiinflammatory,  
CC gastric, immunostimulant and cytostatic activity and can be used  
CC as cadherin-modulators using antisense-therapy or gene-therapy.  
CC These proteins are useful for identifying compounds which bind  
CC or modulate CDNH-1. The proteins, nucleic acids and antibodies  
CC are useful for treating a subject with a disorder characterised  
CC by aberrant or unwanted cadherin protein or nucleic acid activity.  
CC These disorders include central nervous system (e.g. Alzheimer's  
CC disease, Parkinson's disease, multiple sclerosis), cardiovascular,  
CC (e.g. atherosclerosis, angiogenesis), musculoskeletal (e.g. ataxia,  
CC myotonia) gastrointestinal (e.g. gastritis, insulin dependent  
CC diabetes mellitus, Crohn's disease) (e.g. inflammatory and  
CC immune system (e.g. asthma, rheumatoid arthritis, lupus) or cell  
CC proliferation disorders (e.g. cancers, leukaemia) and many other  
CC diseases/disorders listed in the specification.

XX  
SQ Sequence 830 AA;

Query Match 20.3%; Score 877.5; DB 22; Length 830;  
Best Local Similarity 30.5%; Pred. No. 3.7e-62;  
Matches 245; Conservative 141; Mismatches 353; Indels 63; Gaps 22;

Qy 62 LTGETD---NIFVIER-GLLYNRLDRETRTHNLQVAALDANGIIVBGPITIEV 116  
Db LSGDNTADQNTFAVDTSGFLVATRTLDREKAEYQLOVTFLESDGRILGWPLVIVHV 118  
Qy 117 KDINDNPTFLQSYEGSVRONSRGPFLLVYNATDDDPATNGQLYQIVQIOLPMINN 176  
Db KDENDQVPQFSQAIYRAQLSQGTGRGVPFLFLEASDGDGAPGTANSDLRFLHSQSPQPL 178  
Qy 177 VMYQIINKTKAISITREGSQELPAKNPSYNLVSKDMGQSGENSEFSFDTSDIIVTE 236  
Db PDMQLDPLHGLGALALSPGSSLDHALTEYTLQLLVKMDGMDQSPGSHQAIAT-VEISIVE 237  
Qy 237 NIWKAPKPVENVSTDPHIKITQVRWDPGAQYSLVDKELPRFPFSDIQEDGIYVTO 296  
Db NSWAPLEPVHLAENLVVYPHSIAQVHWSGGDVHYQL--ESQPPGPDVDTEGLHWTM 294  
Qy 297 PLDREEDKAYVYAVAKDEYKPLSYPLEIHVKVDINDNPTCPSPVYVEVOENRKL 356  
Db ELDRQAQYQLQVRAQNSHGEDIAEPLQLQVVMENDENDNAPVCSPHDPTVNIPELSPG 354  
Qy 357 NSICTLTAHDREENTANSFLNYRIVEQTKPLMDG-LFLIQTAYAGMLQL-----AKOS 409  
Db TEIARLSAEDLDAPGSPNSHIVYLLSPEPEGAENKAFELDPTSGSVTLCTAPLHAGQS 414  
Qy 410 LKQDTPQYNTLIEV--SDKDFKLCFQINVIDINDQIPFEKSDYGNLFLAEDTNGS 467  
Db LVLAVLAGESSGLSTCEVTVMVTDVNNHAFEFINSQIGVTLPEDEVKGA 469  
Qy 468 TILTIQATDAD-EPTGSSKIL-YHIKGDSEGRGVDTDPHTNGYVYIIKKPLDFETAA 525  
Db LVATLMATDADLEP---AFRLMDEAIEBGGDEGIFDLSEWPSDQVHQLRKNLSYEAP 526  
Qy 526 VSNIVFAENPEPLVFGVKNYASFAKFTLLIVTDVNEAPQSFQVQAKVSEDVAIGTKV 585  
Db DKKVWVSVNTEELV-GPGPGPAATATVTLIRVAVPLKLDQSYETSIPVSTPAGSL 585  
Qy 586 GNVTAQDEGLDISYLRGDFRGMKLDHVTGEIFSVAPLD-REAGSPYRQVVA--TEV 642  
Db LTIQSPDMSRTLRFSLVNDESGHLCIKESVGEVHTAQSLQAGPGDTYTVLVEAQDTDK 645

Qy 643 GGSLSVSSEFHLI-----LMDVNDNPPLKADYTKGLFCHPLSAPGSLIFEATDDQH 696  
Db PGLSTSATVVIHFLKASVPFALTLSAGPSR-----HLCTPRQDYGVVYSGVSEDPDL 697  
Qy 697 LFRGPHTFSLG-SGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRI-NDGGRPPL 754  
Db ANRNGPYSFALGNPTVQRDRLOPLNDSHAYLTALHWPVEGYPVVPVHHHTHMQL 757  
Qy 755 EGVISLPTVFCSC-VEGSCFRPAGHOTGIPVGMAGILLTLLVIGIILAVVFI--- 810  
Db Q-----VKVIVCRCNVEGQMKRKGMPKLSAVGVLLGTAAIGFLLILVFTHLALA 813  
Qy 811 KDKGKDNVESQAQASEVKPLRS 832  
Db RKDL-----QPADSVPLKA 828

#### RESULT 8

AA70741  
ID AAY70741 standard; protein; 906 AA.

XX  
AC AAY70741;

XX  
DT 24-JUL-2000 (first entry)

XX  
DE Human N-cadherin.

XX  
KW Human N-cadherin; Wnt antagonist; contraceptive; contraceptive vaccine;  
KW oocyte development; female primate contraception; oocyte viability;  
KW monoclonal antibody; Wnt signalling.

OS Homo sapiens.

XX  
FH Key Location/Qualifiers  
FT Domain 160..724  
FT /label= Extracellular\_domain  
FT Domain 747..906  
FT /label= Cytoplasmic\_domain

XX  
PN WO200021555-A1.

XX  
PD 20-APR-2000.

XX  
PF 13-OCT-1999; 99WO-US23640.

XX  
PR 15-OCT-1998; 98US-0104355.

XX  
PA (HARD ) HARVARD COLLEGE.

XX  
PI McMahon AP, Parr BA, Vaino S;

XX  
DR WPI; 2000-317845/27.

XX  
PT Contraceptive composition for inhibiting oocyte development in a female  
XX primate comprises a Wnt polypeptide antagonist -

XX  
PS Example 3; Page 27-28; 57pp; English.

XX  
CC The patent discloses a method of female primate contraception comprising  
CC administering an antagonist of a Wnt polypeptide, inhibiting oocyte  
CC development. Wnt polypeptides are useful for promotive maturation of an  
CC immature oocyte. Wnt polypeptides are also useful for increasing the  
CC number of mature oocytes and to enhance oocyte viability. Soluble  
CC fragments of Wnt polypeptides have the ability to inhibit Wnt signalling,  
CC e.g., by blocking binding of a naturally-occurring Wnt protein to its  
CC receptor. They may be used to generate monoclonal antibodies which can  
CC inhibit oocyte development. The present sequence is the human N-cadherin.  
CC Signal transduction by beta-catenin is mediated by binding to the  
CC cytoplasmic domain of a cadherin. Dominant negative mutants of N-cadherin  
CC which inhibit Wnt-1 class signals can also be used as Wnt antagonists.

XX  
SQ Sequence 906 AA;

Query Match	16.2%;	Score 699.5;	DB 21;	Length 906;
Best Local Similarity	27.4%;	Pred. No. 1.3e-47;		
Matches 210;	Conservative 122;	Mismatches 331;	Indels 103;	Gaps 21;
QY	117	KDINDNRPTFLOSKYEGSVQRNRPGLFVYVATDLDLDDPATNGQLYQIVIQPLMINN	176	
DB	47	KDVHEGQP-LLNVKF-----SNCNGKRKQVYESSEPADFKVDEGMYI--AVRSFPLSSE	98	
QY	177	VWYFOINNKTAISLTREGSQELNPAKNPSYLVISVKMGQSGENSEFSFDTTSVDIIV--	234	
DB	99	HAKFLIY-----AQDKETQKQVAVKLSLKPT--LTEESVKESAEEVIFVP	144	
QY	235	-----TENIWKAPKPVEMVNSTDPHPKIKITQVRWN-----DPCAQ	270	
DB	145	RQFSKSHGLQKQKRDWVIP-PINLPENSRGPFQELVRIKSDRDKNLSLRYSVTGPGA-	202	
QY	271	YSLVDREKLPFRFPSIDQ-EGDIVVTPQLDREKDAVFAVAKDEYKGLPSYPLETHVK	329	
DB	203	-----DQPTGIFINPISGQLSVTKPLDREQLARHLRAHVDINGNQVENPIDIVIN	256	
QY	330	VKDINDNRPTCPSPVTVFVEQNERLGNISGTLTAHDEENTANSFLNRYRIVEOTKLP	389	
DB	257	VIDMNDNRPEFLHQVWNGTVPESKPGCTYVMTVAIDADDPNALGNMLRYRIVSQAPSTP	316	
QY	390	MDGLFLIQTAYAGMLQAKQSLKKQDTPQYNLTIEVSKD-----FKTLFCFQVINIDIN	443	
DB	317	SPNFTINNETGDIITVAAGLDREKVQOYTLIIQATDMEGNPTVGLSNTATAVITVDVN	376	
QY	444	DOPIPI-EKSDYGNLTAEINTGSTLTITQATDADDPFGSSKILYHIKIGSEGRGV	502	
DB	377	DNPPEFTAMTYGEV-----PENRVDIIIVANITVDKQDPHTPAWNAVYRISGGDPTGRFAI	433	
QY	503	DTDPHTGTGVITKPLDFTAAVSNIVKAEKPEPLVFGVKYNASSFAKFTLLIVTDVNE	562	
DB	434	QTPDPSNDGLVTVVKPIDFETNRMFVLTVAENQVPLAKGIQHPQSTATVSVTIVDNE	493	
QY	563	APQFSQHVQAKVSEDAVITKGVNVTAKDPEGL---DISYSLRGDPGRWLKIDHVTGEI	619	
DB	494	NPYFAPNPKTIROBEGLHAGTMTTFTAQDPDRYMQOINIRYTKLSDPANWLKIDPVANGQI	553	
QY	620	FSVAPLDREA---GSPYRVQVATEVGGSSLSVSEFHLILMDVNDNPPRLAKDYGLF	675	
DB	554	TTTAVLDRESPVNKNINYNATFELASDNGIPPMSTGTGLQIYLLDINDNAPQVL-----	606	
QY	676	FCHPLSA-----PGSLIFEATDDQHLFRGPH-FTFSLGSGSLQNDWEVSKINGTHAR	727	
DB	607	---PQEAETCTPDPSINITALDYDIDPNAGPFAFDPLPSPTIKRNTWITRLNGDFAQ	663	
QY	728	LSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSC-VEGSCFRPAGHOTGIPTV-	785	
DB	664	LNLKIKFLEAGIYEVPIIITDSGNPPKSNISILRVKVCQDSNGDC-----TDVDRIV	716	
QY	786	-GMVAGILLTLLVIGIILAVFIRIKKDKGKDNVESQAASEVKP	829	
DB	717	GAGLGTGAILLILLCIILLILVLMFVVMKRRDKERQAKOLLIDP	762	
RESULT 9				
ABG14316				
ID	ABG14316 standard; Protein; 906 AA.			
XX	ABG14316;			
XX	18-FEB-2002 (first entry)			
DT	Novel human diagnostic protein #14307.			
XX	Human; chromosome mapping; gene mapping; gene therapy; forensic;			
KW	food supplement; medical imaging; diagnostic; genetic disorder.			
XX	Homo sapiens.			
OS	WO200175067-A2.			
PN				

XX	11-OCT-2001.				
PD	30-MAR-2001; 2001WO-US08631.				
XX	31-MAR-2000; 2000US-0540217.				
PR	23-AUG-2000; 2000US-0649167.				
XX	(HYSE-) HYSEQ INC.				
XX	Dmanac RT, Liu C, Tang YT;				
PI	WPI; 2001-639362/73.				
XX	N-PSDB; AAS78503.				
DR	New isolated polynucleotide and encoded polypeptides, useful in				
XX	diagnostics, forensics, gene mapping, identification of mutations				
PT	responsible for genetic disorders or other traits and to assess				
PT	biodiversity				
XX	Claim 20; SEQ ID No 44675; 103pp; English.				
PS	The invention relates to isolated polynucleotide (I) and				
XX	polypeptide (II) sequences. (I) is useful as hybridisation probes,				
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome				
CC	and gene mapping, and in recombinant production of (II). The				
CC	polynucleotides are also used in diagnostics as expressed sequence tags				
CC	for identifying expressed genes. (I) is useful in gene therapy techniques				
CC	to restore normal activity of (II) or to treat disease states involving				
CC	(II). (II) is useful for generating antibodies against it, detecting or				
CC	quantitating a polypeptide in tissue, as molecular weight markers and as				
CC	a food supplement. (II) and its binding partners are useful in medical				
CC	imaging of sites expressing (II). (I) and (II) are useful for treating				
CC	disorders involving aberrant protein expression or biological activity.				
CC	The polypeptide and polynucleotide sequences have applications in				
CC	diagnostics, forensics, gene mapping, identification of mutations				
CC	responsible for genetic disorders or other traits to assess biodiversity				
CC	and to produce other types of data and products dependent on DNA and				
CC	amino acid sequences. ABG00010-ABG30377 represent novel human				
CC	diagnostic amino acid sequences of the invention.				
CC	Note: The sequence data for this patent did not appear in the printed				
CC	specification, but was obtained in electronic format directly from WIPO				
CC	at ftp.wipo.int/pub/published_pct_sequences.				
XX					
SQ	Sequence 906 AA;				
Query Match		16.2%;	Score 699.5;	DB 22;	Length 906;
Best Local Similarity		27.4%;	Pred. No. 1.3e-47;		
Matches 210;		Conservative 122;	Mismatches 331;	Indels 103;	Gaps 21;
QY	117	KDINDNRPTFLOSKYEGSVQRNRPGLFVYVATDLDLDDPATNGQLYQIVIQPLMINN	176		
DB	47	KDVHEGQP-LLNVKF-----SNCNGKRKQVYESSEPADFKVDEGMYI--AVRSFPLSSE	98		
QY	177	VWYFOINNKTAISLTREGSQELNPAKNPSYLVISVKMGQSGENSEFSFDTTSVDIIV--	234		
DB	99	HAKFLIY-----AQDKETQKQVAVKLSLKPT--LTEESVKESAEEVIFVP	144		
QY	235	-----TENIWKAPKPVEMVNSTDPHPKIKITQVRWN-----DPCAQ	270		
DB	145	RQFSKSHGLQKQKRDWVIP-PINLPENSRGPFQELVRIKSDRDKNLSLRYSVTGPGA-	202		
QY	271	YSLVDREKLPFRFPSIDQ-EGDIVVTPQLDREKDAVFAVAKDEYKGLPSYPLETHVK	329		
DB	203	-----DQPTGIFINPISGQLSVTKPLDREQLARHLRAHVDINGNQVENPIDIVIN	256		
QY	330	VKDINDNRPTCPSPVTVFVEQNERLGNISGTLTAHDEENTANSFLNRYRIVEOTKLP	389		
DB	257	VIDMNDNRPEFLHQVWNGTVPESKPGCTYVMTVAIDADDPNALGNMLRYRIVSQAPSTP	316		
QY	390	MDGLFLIQTAYAGMLQAKQSLKKQDTPQYNLTIEVSKD-----FKTLFCFQVINIDIN	443		
DB	317	SPNFTINNETGDIITVAAGLDREKVQOYTLIIQATDMEGNPTVGLSNTATAVITVDVN	376		











QY 817 D 817  
Db 762 E 762

RESULT 15  
AAR63533  
ID AAR63533 standard; Protein; 896 AA.  
XX  
XX AAR63533;  
XX  
XX 07-JUN-1995 (first entry)  
XX  
XX Human HT-1376 cell-derived Leukocyte Chemotactic Factor.  
XX  
XX rHT-LCF; HT-1376 cell derived leukocyte chemotactic factor;  
KW immunopotentiator; monocyte and macrophage migration factor;  
KW wound healing.  
XX  
XX Homo sapiens (cell line HT-1376).  
XX  
XX Key Location/Qualifiers  
FH Peptide 1..30  
FT /label= signal\_peptide  
FT Protein 31..896  
FT /label= rHT-LCF  
XX  
XX W09421809-A.  
XX  
XX 29-SEP-1994.  
XX  
XX 11-MAR-1994; 94WO-JP00397.  
XX  
XX 15-MAR-1993; 93JP-0054349.  
XX  
XX (CYTO-) INST CYTOSIGNAL RES INC.  
XX  
XX Kawamura K, Watanabe K;  
XX WPI; 1994-317029/39.  
XX N-PSDB; AAQ72597.  
XX  
XX Monocyte and macrophage migration factor protein - is  
PT immuno-potentiator and vulnerary and is isolated from culture of  
PT HT-1376 cell line by recombinant methods  
XX  
XX Claim 5; Page 34-41; 68pp; Japanese.  
XX  
XX Sequences coding for proteins having monocyte and macrophage  
CC migration factor activity were isolated from a cDNA library  
CC prepared from human HT-1376 cells (ATCC CRL-1472). The proteins  
CC encoded by the isolated cDNA clones are useful as  
CC immunopotentiators for treatment of infections, immune  
CC deficiency diseases and cancer. The proteins are also useful in  
CC wound healing preparations and for treatment of skin diseases.  
XX  
XX Sequence 896 AA;

Query Match 15.5%; Score 671.5; DB 15; Length 896;  
Best Local Similarity 30.8%; Pred. No. 2.5e-45;  
Matches 187; Conservative 106; Mismatches 263; Indels 51; Gaps 22;

QY 239 WKAKPKVEMVNSTDPHPKIKITQVRWNPDAQYSL-----VDKEKLPRFPESIDQEG 290  
Db 137 W-APIPCSMQENSLGPFPLFLOQVE-SDAQNYTVFYSISGRGVDXEPLNLFYIERD-TG 193  
QY 291 DIYVTPQLDREKDAY--VFYAVAKDEYKPLSYPLETHVVKVDINDNPPCTCPSPVTVE 348  
Db 194 NLCTRPVDREYDVFOLIAVASTADGYSADL--PLPLPIRVDENDNHPVFEAIYNFE 251  
QY 349 VQENERIGNSICTLTAHDEDENTANSFLNRYVEQTPKLPMDGLFLIOTYAGMLQALQ 408  
Db 252 VLESSRPCTTGVGCATDRDEPDTMTRLKYSILQOTPRSP--GLFSVHPSTGVITTVSH 309

Search completed: April 2, 2003, 15:52:23  
Job time : 67.6849 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 2, 2003, 15:51:14 ; Search time 35.1416 Seconds  
(without alignments)  
696.608 Million cell updates/sec

Title: US-09-079-678-178  
Perfect score: 4321  
Sequence: 1 MILOAHLHSLCLMLLYLATG.....DKGKNVESAQAASEVKPLRS 832

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA: \*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep: \*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep: \*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep: \*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep: \*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep: \*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4321	100.0	832	1	US-08-431-560-1
2	4321	100.0	832	1	US-08-463-345-1
3	696	16.1	913	2	US-08-474-067-6
4	696	16.1	913	2	US-08-474-068A-6
5	696	16.1	913	2	US-08-472-481-5
6	684.5	15.8	916	1	US-08-188-228-48
7	684.5	15.8	916	1	US-08-332-643-42
8	684.5	15.8	916	1	US-08-332-638-48
9	657.5	15.2	713	1	US-08-188-228-62
10	657.5	15.2	713	1	US-08-332-643-56
11	657.5	15.2	713	1	US-08-332-638-62
12	650.5	15.1	555	2	US-08-453-702A-98
13	634	14.7	837	2	US-08-474-067-7
14	634	14.7	837	2	US-08-474-068A-7
15	634	14.7	837	2	US-08-472-481-6
16	633	14.6	556	1	US-07-998-003A-98
17	633	14.6	556	1	US-08-453-274B-98
18	633	14.6	556	1	US-08-453-695A-98
19	633	14.6	556	1	US-08-268-161A-98
20	633	14.6	556	4	US-09-099-639-98
21	630.5	14.6	712	2	US-08-474-067-2
22	630.5	14.6	712	2	US-08-474-067-5
23	630.5	14.6	712	2	US-08-474-068A-2
24	630.5	14.6	712	2	US-08-474-068A-5
25	630.5	14.6	712	2	US-08-472-481-2
26	630.5	14.6	717	2	US-08-474-067-4
27	630.5	14.6	717	2	US-08-474-068A-4

28	630.5	14.6	717	2	US-08-472-481-4	Sequence 4, Appl
29	608	14.1	884	2	US-08-474-067-8	Sequence 8, Appl
30	608	14.1	884	2	US-08-474-068A-8	Sequence 8, Appl
31	608	14.1	884	2	US-08-472-481-7	Sequence 7, Appl
32	593	13.7	878	1	US-08-237-919-2	Sequence 2, Appl
33	593	13.7	878	4	US-08-732-429-2	Sequence 2, Appl
34	593	13.7	878	4	US-09-798-267-3	Sequence 3, Appl
35	593	13.7	878	4	US-09-798-267-3	Sequence 3, Appl
36	593	13.7	878	5	PCT-US95-05518-2	Sequence 25, Appl
37	592.5	13.7	117	4	US-09-187-859-25	Sequence 2, Appl
38	574.5	13.3	796	2	US-08-738-349-2	Sequence 4, Appl
39	560.5	13.0	796	2	US-08-738-349-4	Sequence 4, Appl
40	559	12.9	822	2	US-08-474-067-9	Sequence 9, Appl
41	559	12.9	822	2	US-08-474-068A-9	Sequence 9, Appl
42	559	12.9	822	2	US-08-472-481-8	Sequence 8, Appl
43	555	12.8	106	1	US-08-326-117B-6	Sequence 6, Appl
44	555	12.8	106	3	US-08-982-129-6	Sequence 6, Appl
45	554	12.8	106	4	US-09-178-176B-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1  
US-08-431-560-1  
; Sequence 1, Application US/08431560  
; Patent No. 5620855  
; GENERAL INFORMATION:  
; APPLICANT: Anne H. Dantzig, et al.  
; TITLE OF INVENTION: Mammalian Influx Peptide  
; TITLE OF INVENTION: Transporter  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Eli Lilly and Company  
; STREET: Lilly Corporate Center  
; CITY: Indianapolis  
; STATE: IN  
; COUNTRY: U.S.A.  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; Medium TYPE: Diskette, 3.50 inch, 1.0 Mb storage  
; COMPUTER: Macintosh  
; OPERATING SYSTEM: Macintosh  
; SOFTWARE: Microsoft Word  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/431,560  
; FILING DATE: 01-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/013,462  
; FILING DATE: 04-FEB-1993  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 832 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-431-560-1

Query Match	100.0%;	Score 4321;	DB 1;	Length 832;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 832;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MILOAHLHSLCLMLLYLATGYGOEGKFSGLPKDWTESIYEGQPSOLIIFQFKANPAVTF	60	
Db	1	MILOAHLHSLCLMLLYLATGYGOEGKFSGLPKDWTESIYEGQPSOLIIFQFKANPAVTF	60	
QY	61	ELTGETDNFVIERGLLYNNRDLRETRSTHNLQVAALDANGIIYEGVPPIETVKDIN	120	
Db	61	ELTGETDNFVIERGLLYNNRDLRETRSTHNLQVAALDANGIIYEGVPPIETVKDIN	120	
QY	121	DNRPTEFLQSKYEGSVQNSRPGKPFLLYVNATDLDPATPNQLYYQIVITOLPNNMYVF	180	
Db	121	DNRPTEFLQSKYEGSVQNSRPGKPFLLYVNATDLDPATPNQLYYQIVITOLPNNMYVF	180	

Db 121 DNRPTFLOSKYEGSVRONSRPGKPFYLVNATDLDPATPNQQLYQYIVIQPLMNNVMYF 180  
QY 181 QINNKTGAISLTREGSQELNPAKNPSYLVSKDMGGQSENFSFSDTSDTSYDIIIVTENIWK 240  
Db 181 QINNKTGAISLTREGSQELNPAKNPSYLVSKDMGGQSENFSFSDTSDTSYDIIIVTENIWK 240  
QY 241 APKPVEVNSTDPHPKTIQVRWNDGCAQYSLVDKEKLPFRPFSDIQEGDIYVTPQLDR 300  
Db 241 APKPVEVNSTDPHPKTIQVRWNDGCAQYSLVDKEKLPFRPFSDIQEGDIYVTPQLDR 300  
QY 301 EEDKAYVYAVADEYKGLPSYPLEIHVKVDINDNPPCTCPSPVTVEVOENRGLNSIG 360  
Db 301 EEDKAYVYAVADEYKGLPSYPLEIHVKVDINDNPPCTCPSPVTVEVOENRGLNSIG 360  
QY 361 TLTAHDEENTANSFLNYRIVEQTPKLPMDGLFLIQTAYAGMLQAKQSLKKQDTPQYNL 420  
Db 361 TLTAHDEENTANSFLNYRIVEQTPKLPMDGLFLIQTAYAGMLQAKQSLKKQDTPQYNL 420  
QY 421 TIEVSDKDFKTLCFVQINVIDINDQIPIFEKSDYGNLTAEEDTNGSTILTIOATDAEP 480  
Db 421 TIEVSDKDFKTLCFVQINVIDINDQIPIFEKSDYGNLTAEEDTNGSTILTIOATDAEP 480  
QY 481 FTGSSKILYHIKGDSEGRGLVDTDPHTNTGYVLIKKPLDFETAASNIVFKAENPEPLV 540  
Db 481 FTGSSKILYHIKGDSEGRGLVDTDPHTNTGYVLIKKPLDFETAASNIVFKAENPEPLV 540  
QY 541 FGKYNASSFAKFTLIIVTDVNEAPQFSQHVFOAKVSEDVAIGTKVGNVTAKDPEGLDISY 600  
Db 541 FGKYNASSFAKFTLIIVTDVNEAPQFSQHVFOAKVSEDVAIGTKVGNVTAKDPEGLDISY 600  
QY 601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVADEVGSSLSVSEPHLIIMDV 660  
Db 601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVADEVGSSLSVSEPHLIIMDV 660  
QY 661 NDNPRLAKDYTGGLFFCHPLSAPGSLIFEATDDOHLFRGPHFTFSLGSGSLQNDWEVSK 720  
Db 661 NDNPRLAKDYTGGLFFCHPLSAPGSLIFEATDDOHLFRGPHFTFSLGSGSLQNDWEVSK 720  
QY 721 INGHARLSTRHTDFEERAYVVLIRINDGRRPPLLEGIVSLPVTFSCVEGSCFRPAGHOT 780  
Db 721 INGHARLSTRHTDFEERAYVVLIRINDGRRPPLLEGIVSLPVTFSCVEGSCFRPAGHOT 780  
QY 781 GIPTVGMVAGILLTLLVIGIILAVFIRIKKDKGKDNVESQAQSEVKPLRS 832  
Db 781 GIPTVGMVAGILLTLLVIGIILAVFIRIKKDKGKDNVESQAQSEVKPLRS 832

RESULT 2

US-08-463-345-1  
; Sequence 1, Application US/08463345  
; Patent No. 5710018  
; GENERAL INFORMATION:  
; APPLICANT: Anne H. Dantzig, et al.  
; TITLE OF INVENTION: Mammalian Influx Peptide  
; TITLE OF INVENTION: Transporter  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Eli Lilly and Company  
; STREET: Lilly Corporate Center  
; CITY: Indianapolis  
; STATE: IN  
; COUNTRY: U.S.A.  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.0 Mb storage  
; OPERATING SYSTEM: Macintosh  
; SOFTWARE: Microsoft Word  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/463,345  
; FILING DATE: 05-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/013,462  
; FILING DATE: 04-FEB-1993  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 832 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-463-345-1

Query Match 100.0%; Score 4321; DB 1; Length 832;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 832; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MILQAHLHSCLLMLYLATGYGQEGKFSGLPKMTFSIYEGQPSQIIFQKANPPAVTF 60  
Db 1 MILQAHLHSCLLMLYLATGYGQEGKFSGLPKMTFSIYEGQPSQIIFQKANPPAVTF 60  
QY 61 ELTGETDNIFVIREGGLLYNRRALDRETRSTHNLQVAALDANGIIIVEGPPVITIEVKDIN 120  
Db 61 ELTGETDNIFVIREGGLLYNRRALDRETRSTHNLQVAALDANGIIIVEGPPVITIEVKDIN 120  
QY 121 DNRPTFLOSKYEGSVRONSRPGKPFYLVNATDLDPATPNQQLYQYIVIQPLMNNVMYF 180  
Db 121 DNRPTFLOSKYEGSVRONSRPGKPFYLVNATDLDPATPNQQLYQYIVIQPLMNNVMYF 180  
QY 181 QINNKTGAISLTREGSQELNPAKNPSYLVSKDMGGQSENFSFSDTSDTSYDIIIVTENIWK 240  
Db 181 QINNKTGAISLTREGSQELNPAKNPSYLVSKDMGGQSENFSFSDTSDTSYDIIIVTENIWK 240  
QY 241 APKPVEVNSTDPHPKTIQVRWNDGCAQYSLVDKEKLPFRPFSDIQEGDIYVTPQLDR 300  
Db 241 APKPVEVNSTDPHPKTIQVRWNDGCAQYSLVDKEKLPFRPFSDIQEGDIYVTPQLDR 300  
QY 301 EEDKAYVYAVADEYKGLPSYPLEIHVKVDINDNPPCTCPSPVTVEVOENRGLNSIG 360  
Db 301 EEDKAYVYAVADEYKGLPSYPLEIHVKVDINDNPPCTCPSPVTVEVOENRGLNSIG 360  
QY 361 TLTAHDEENTANSFLNYRIVEQTPKLPMDGLFLIQTAYAGMLQAKQSLKKQDTPQYNL 420  
Db 361 TLTAHDEENTANSFLNYRIVEQTPKLPMDGLFLIQTAYAGMLQAKQSLKKQDTPQYNL 420  
QY 421 TIEVSDKDFKTLCFVQINVIDINDQIPIFEKSDYGNLTAEEDTNGSTILTIOATDAEP 480  
Db 421 TIEVSDKDFKTLCFVQINVIDINDQIPIFEKSDYGNLTAEEDTNGSTILTIOATDAEP 480  
QY 481 FTGSSKILYHIKGDSEGRGLVDTDPHTNTGYVLIKKPLDFETAASNIVFKAENPEPLV 540  
Db 481 FTGSSKILYHIKGDSEGRGLVDTDPHTNTGYVLIKKPLDFETAASNIVFKAENPEPLV 540  
QY 541 FGKYNASSFAKFTLIIVTDVNEAPQFSQHVFOAKVSEDVAIGTKVGNVTAKDPEGLDISY 600  
Db 541 FGKYNASSFAKFTLIIVTDVNEAPQFSQHVFOAKVSEDVAIGTKVGNVTAKDPEGLDISY 600  
QY 601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVADEVGSSLSVSEPHLIIMDV 660  
Db 601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVADEVGSSLSVSEPHLIIMDV 660  
QY 661 NDNPRLAKDYTGGLFFCHPLSAPGSLIFEATDDOHLFRGPHFTFSLGSGSLQNDWEVSK 720  
Db 661 NDNPRLAKDYTGGLFFCHPLSAPGSLIFEATDDOHLFRGPHFTFSLGSGSLQNDWEVSK 720  
QY 721 INGHARLSTRHTDFEERAYVVLIRINDGRRPPLLEGIVSLPVTFSCVEGSCFRPAGHOT 780  
Db 721 INGHARLSTRHTDFEERAYVVLIRINDGRRPPLLEGIVSLPVTFSCVEGSCFRPAGHOT 780  
QY 781 GIPTVGMVAGILLTLLVIGIILAVFIRIKKDKGKDNVESQAQSEVKPLRS 832  
Db 781 GIPTVGMVAGILLTLLVIGIILAVFIRIKKDKGKDNVESQAQSEVKPLRS 832

RESULT 3  
US-08-474-067-6

```

; Sequence 6, Application US/08474067
; Patent No. 5811518
; GENERAL INFORMATION:
; APPLICANT: Ranscht, Barbara
; TITLE OF INVENTION: T-Cadherin Adhesion Molecule
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,067
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/213,361
; FILING DATE: 14-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/607,293
; FILING DATE: 30-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1682
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-474-067-6

```

```

Query Match 16.1%; Score 696; DB 2; Length 913;
Best Local Similarity 26.9%; Pred. No. 6.9e-53;
Matches 209; Conservative 117; Mismatches 328; Indels 124; Gaps 22;

QY 126 FLSKVEGSRONRCKPPLVYNATDLPATPNGQLYQIVQLPMINNMYEQIN-- 183
DB 41 FPDVHSAVSVSRVHGGQPLLNVRFQSCDE-----NRKIYFGSSEP 81

QY 184 -----NKTGAISLTREGSQELNPAKNPSYNLVISVKDMGQGS-----ENFSFDTT 228
DB 82 EDFRVGEGVVYAERSQLSAEPE-----FVVSARDKETQEWQMKVLTPEPAFTGAS 136

QY 229 SVDIIVTENI-----WKAPKPVEMVENSSTDPHPKIKITQVRWNPDA 269
DB 137 EKOKKIEDIFFWQYKDSHLKROKRDWVIP--PINLPENSARGPPQELVRIR----- 189

QY 270 QYSLVDKELPRPPEST-----DQ-----EGDIYVTPDLDEKDAVYFVAVAK 313
DB 190 -----SDRDKSLSLRYSTVGCGADQPPGTGFIINPISQLSVTKPLDREQIASFLHRAHV 245

QY 314 DEXGKPLSYPLEIHVKYKIDNDPPTCPSPVTVFEVGENERLGNISGLTAAHDEENTA 373
DB 246 DVNGQVENPIDIVINDNDRPELHQWNGTVPEGSKPGTYVMTVTAIDADDNAQ 305

QY 374 NSFNLRYIVQETKPLMDGLFLIQTYAGMLQAKQSLKKQDTPQYNLTIEVSKD----- 428
DB 306 NGMLRYRLSQAPSSPNMFTINNETGDIITVAAGLDREKVQYQYTLIIQATDMEGNPTY 365

QY 429 -FXTLCFQVQINVDINDQIPF-EKSDYGNLTAEADNIGSTILTIQATDADEFTGSSK 486
DB 366 GLSNATAVITVTDVNDNPPFTAMTEYGEV---PENRVDVIVANLTVTDKDPQHPAWN 422

```

```

QY 487 ILXHIKGDSEGRGLGVDTPHTNTGYVVIKKPLDFETAAYSNIVFKAENPEPLVGVKYN 546
DB 423 ARYQMTGDDPTGGFTILTDPSNNDCLVTVVVKPIDFETNRMFVLTVAAENQVPLAKGIQHP 482
QY 547 ASSFAKFTLIVTDVNEAPQFSQHVFAKVSQEDVAIGTKVGNVTAKDP-----EGLDISYSL 602
DB 483 PQSTATVSTIVDVNESYFVPNPKLVQREBGLLAGSMLTFTTARDPDYMQOOTSLSYSK 542
QY 503 RGDTRGHLKIDHVTGEIFSVAPLDREA----GSPYRVQVVATEVGGSSLSVSEFHLILM 658
DB 543 LSDPANWLKIDPVNGQITTTAVLDRESIYVQNNMYNATFLASDNGIIPMSGTGLQIYLL 602
QY 659 DVNDNPPRL-AKDYTGFLFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSG--SLQND 715
DB 503 DINDNAPQVNPKEAT---TCETLQ-PNAINITAVDPDIDPNAGP-FAFELPDSPPSIKRN 657
QY 716 WEVSKINGTHARLSTRHTDFEERAYVVLIRINDGRRPPLLEGIVSLVPTFCSC-VEGSCFR 774
DB 658 WTVIRISGDHAQLSLRIRFLEAGIYDVPITVDSGNPHASSTSVLKVKVCCODINGDC-- 715
QY 775 PAGHOTGIPTV---GMAVGILLTLLVIGIILAVVFIIRIKKDKGKNVESQAQASEVXP 829
DB 716 -----TDVDRIVGAGLGTGAIITAILLIIILVLMFVVMKRRDKERQAKQLLIDP 768

```

```

RESULT 4
US-08-474-068A-6
; Sequence 6, Application US/08474068A
; Patent No. 5837525
; GENERAL INFORMATION:
; APPLICANT: Ranscht, Barbara
; TITLE OF INVENTION: T-Cadherin Adhesion Molecule
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,068A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/213,361
; FILING DATE: 14-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/607,293
; FILING DATE: 30-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1683
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-474-068A-6

```

```

Query Match 16.1%; Score 696; DB 2; Length 913;
Best Local Similarity 26.9%; Pred. No. 6.9e-53;
Matches 209; Conservative 117; Mismatches 328; Indels 124; Gaps 22;

```







```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-332-643-42

```

Query Match	15.8%	Score	684.5	DB 1	Length	916
Best Local Similarity	29.3%	Pred. No.	7.3e-52			
Matches 176	Conservative	106	Mismatches	282		
			Indels	31	Gaps	12

FILING DATE: 01 NOV 1994  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/872,643  
 FILING DATE: 17 APR 1992  
 APPLICATION NUMBER: US/08/049,460  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: No. 5646250and, Greta E.  
 REGISTRATION NUMBER: 35,902  
 REFERENCE/DOCKET NUMBER: 31340  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (312) 474-6300  
 TELEFAX: (312) 474-0448  
 TELEX: 25-3856  
 INFORMATION FOR SEQ ID NO: 48:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 916 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-332-638-48

Query Match	15.8%	Score 684.5;	DB 1;	Length 916;
Best local similarity	20.3%			
Pred NO	7	3e-52.		

12;

QY	239	WKAPKPVEMVNSTDPHPKIKTQVRW---NDPQAQVSL--VDKEKLPURFPFFSID-QBGDI	292
Db	171	WVIP-PINVPENSRGPPQOOLVRIKSDKNDIPIRYSITGVGADQPMPEVFSINSGRM	229
QY	293	YVTPQLDREKDAYFYAVAKDEYKPLSVLEIHWKVDINDNPCTCPSVTVVEQGN	352
Db	230	YVTRPMDEEBASHULRAHVAIDMNGKNVENPDLIYIVIDMNDHPEFTINQVYNSVDGG	289
QY	353	ERLNGSIGTLTAHREDEENTASFELNRYRVEQTPKLPMDGLFIQTYAGMLQAKSLKK	412
Db	290	SKPGTYVMTITANDADDSHTANGWRYRIVTQTPQSPQNMFETINSETGDIVTVAAQNDR	349
QY	413	QDPQPNLTIIEVSDKD-----FKTLCEVQIINVIDNDQIPIFEKSDYCNLTLAEDTNG	466
Db	350	EKVQOQYTVIVQATDMGNLNYGLNSTATALLVTVDNDNPSEFTASTFAG--EVPENSVE	407
QY	467	STILTQATDADPEFTGSSKILYHIITKDSBGRGLGVDTDPHNTGVVITKPLDPEATAV	526
Db	408	TVVANILTVMDRQDPHSPNNNAVYRIISGDPSGHFSVTRTDPVTENEGMAVTVVKVADYELNRA	467
QY	527	SNTVFKAENPEPTVFGVKYNASSFAKFTLITVDVNEAPOSFOHVFOAKYSEDAVIGTKVG	586

## RESULT 8

US-08-332-638-48  
: Sequence 48: Application US/08332638

; sequence 40; application 03/08332038  
; Patent No. 5646250

GENERAL INFORMATION:

; APPLICANT: Suzuki, Shintaro

; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS

; NUMBER OF SEQUENCES: 62

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &

; ADDRESSEE: Borun

STREET: 6300

CITY: Chicago

STATE: Illinois  
COUNTRY: USA

; COUNTRY: USA  
; ZIP: 60606; ZIP: 60606  
; COMPUTER READABLE FORM.

```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS

OPERATING SYSTEM: PC-DOS/MS-DOS

10

```
RESULT 9
US-08-188-228-62
; Sequence 62, Application US/08188228
; Patent No. 5597725
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Borun
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: US/08/188,228
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/049,460
; FILING DATE: 19 APR 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,643
; FILING DATE: 17 APR 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5597725and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31340
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 713 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-188-228-62
Query Match 15.2%; Score 657.5; DB 1; Length 713;
Best Local Similarity 28.7%; Pred. No. 1.2e-49;
Matches 199; Conservative 104; Mismatches 285; Indels 105; Gaps 23;
QY 160 NGQLYQIVQLPMNNVWFQINNKTCALSL---TREGSOELNPAKNPSYN-----LVI 211
DB 61 NDLRIEY-----SSPYFKVNSDGLVALRNITAVGKTLFVHARTHAEDMAELIV 112
QY 212 SVKDMGQSEN--SFSDTTSV-----DIIVTENIWKAPKPVEMVENSTDPHPKIKITQVRW 264
DB 113 GGRDIQSLQIDPKFARTSPVPRQKRSIVS-----PILPENORQPPRPDRVGKV-- 162
QY 265 NDFGAQYSLVDKELKLPFPF-----SIDQE-----GDIYVTPQLDREKDAYVF 308
DB 163 -----VDSRDPERSKFLTGKGVQEPKGFIRINENTGVSVTRTLDREVIAVYQL 213
QY 309 YAVAKDEYKPLSPLEIHVKVDKINDNPTCPSPVTFVEQENERLGNISGILTANDRD 368
DB 214 FVETDVGKTLGEPVLEIVIDQNDNRPIFREGPYIGHVMEGSPGTGTVMRMTAFDAD 273
QY 369 EENTANSFLNRYVEOTPKLPMDGLFLIQTYAG-MLQAKQSLKKQDT---POYNLTIEV 424
DB 274 DPATDNALLRYNRQOTPKRPSNMFYIDPKGDIVTVVSPALLDRETLENPKYELIEA 333
QY 425 SDK-----DFKTLCFVQINVIDINDQIPIFEKSDYGNLTIAEDTNIIGTILTIQATDADE 479
DB 334 QDMAGLDVGLTGTATATIMIDDKNDHSPKFTKKEF--QATVEEGAVG-VIVNLTVEDKDD 390
QY 480 PFTGSSKILYHIKGDSEGRGLCVDTDPHTWTGYIIKKPLDFETAAYASNIVFKAENPEPL 539
DB 391 PTTGAWRAAYTIINGNPGQSFETHNTQTNEGMLSVVVKPLDYEISAFHTLLIKYVENEDPL 450
QY 540 VFGVKNASSGFAKFTLIIVTDVNEAPQFSQHVFOAKVSEDVAIGTKVGNVTAKPDEGLD-- 597
DB 451 VPDVSYGSPSTATVHITVLDVNEGPVYPPDMVMTREDDLSVGSVLITVNAATDPSLQHQ 510
QY 598 -ISYLRGDRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQVQVATEVGGSSLSVSE 652
DB 511 TIRYSVKYDKPAGMLNINPINGTVDTTAVLDRESFVNSVVTALFLAIDSGNPATGTGT 570
QY 653 FHLILMDVNDNPPRLAKDYTGLFCHPLSAP-----GSLIFEATDDDOHLERGPHT 704
DB 571 LLITLEDVNDNAP-----FIYPTVAEVCDDAKNLSVILGASDKDLHPNTDP-FK 619
QY 705 FSLGSGSLQND-WEVSKINGTHARLSTRHTDFFERAYVVLIRINDGGRPPLEGIVSLPVT 763
DB 620 FEIHKQAVDPKVKWKISKINNTHALVSLQ-NLNKANYNLPIWVTDSCGPPMTNITDLRVQ 678
QY 764 FCSCVEG--SCFRPAGHOTGIPTVGMVAGILLT 794
DB 679 VCSRNKSKVDCNAAGALRFSLSV-----ILLS 706
RESULT 10
US-08-332-643-56
; Sequence 56, Application US/08332643
; Patent No. 5639634
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
; STREET: Two First National Plaza, 20 South Clark
; STREET: Street
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/332,643
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/872,643
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5639634and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/30795
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 346-5750
; TELEFAX: (312) 984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 713 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-332-643-56
```

```

Query Match      15.2%; Score 657.5; DB 1; Length 713;
Best Local Similarity 28.7%; Pred. No. 1.2e-49;
Matches 199; Conservative 104; Mismatches 285; Indels 105; Gaps 23;

Qy 160 NGQYYQIVQIPLMNNVYFQINKTKAISL---TREGSOELNPAKNPSYN-----LVI 211
Db 61 NDKLRYEV-----SSPYFKVNSDGLVALRNITAVGKTLFVHARTPHAEDEMAELVIV 112
Qy 212 SVKDMGQSEN--SFSDDTSV-----DIIVTENIWKAPKPVEMVENSDDPHIKITOVWR 264
Db 113 GKGDIQSLQDIFKARTSPVPROKRSIVS-----PILIPENQRPDRVGKV-- 162
Qy 265 NDPGAQYSLVDKEKLPRFPF-----SIDQE-----GDIYVTPQLDREKDAYVF 308
Db 163 -----VSDRSPERSKFRLTGKGVQDEPKGFRINENTGSVSVTRTLDREVIAYVOL 213
Qy 309 YAVAKDEYKPLSVLEIHKVVKDINDNPPCTPSPVTVFEVQENRGLNSIGTLTAHRRD 368
Db 214 FVETDVGKTLGEGVPLEVIVDQNDNRPIFREGPYIGHVMEGSPGTGTTVMRTAFDAD 273
Qy 369 EENTANSFLNRYVEQTPKLPMDGLFLIQTAYG--MLQAKOSLKKQDT---POYNLTIEV 424
Db 274 DPATDNALLRYNIRQQTDPKPSNMFYIDPEKGDIVTVVSPALLDRETLNPKYELLIEA 333
Qy 425 SDK-----DFKTLCFVQINVIDINDQIPIFEKSDYGNLTJLAEDTNIGSTILTIQATDAE 479
Db 334 QDMAGLDVGLGTATATIMIDDKNDHSPKFTKKEF--QATVEEGAVG--VIVNLTVEDKDD 390
Qy 480 PFTGSSKILYHIKGDSEGRIGVTDPHNTGYVLIKKPLDFETAANSVIVFKAENPEPL 539
Db 391 PTTGAWRAAYTIINGNPGQSFIEHTNPQTNQNEGLSVKPLDYEISAFHTLLIKYENEDPL 450
Qy 540 VFGVKNASSFAKFTLIIVTDVNEAPQFSQHVFOAKVSDVAIGTKVGNVTAKDPEGLD-- 597
Db 451 VPDVSGPSSTATVHITVLDVNEGVPYPPDMVMYTRQEDLSVGSVLLTVNATDPSLQHQ 510
Qy 598 -ISYSLRGDTRGWLKIDHVTGEIFSAPLDREA----GSPYRVQVWATEVGGSSLSVSE 652
Db 511 TIRSVYKDPAGWLNINPINGTVDTTAVLDRESFPVDSVYTFALFLAIDSNPATGTGT 570
Qy 653 FHLIMDVNDNPPRLAKDYTGFLFCHPLSAP-----GSLIFEATDDDOHLFRGPHFT 704
Db 571 LLITLEDVNDNAP-----FIYPTVAEVCDDAKNLSVILGASDKDLHPNTDP--FK 619
Qy 705 FSLGSGSLOND--WEVSKINGTHARLSTRHTDPEERAYVVLIRINDGGRPPLEGIVSLPVT 763
Db 620 FEIHKQAVDPKVKISKINNTHALVSLIQ--NLNKANYNLPIMWTDGSKPPMTNITDLRVQ 678
Qy 764 FCSCVEG--SCFRPAGHOTGTPVGMAGVILLT 794
Db 679 VCSRSKVDCAAGALRFLSPV-----ILLS 706

RESULT 11
US-08-332-638-62
; Sequence 62, Application US/08332638
; Patent No. 5646250
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Borun
; ADDRESS: Marshall, O'Toole, Gerstein, Murray &
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/332,638
; FILING DATE: 01-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,643
; FILING DATE: 17 APR 1992
; APPLICATION NUMBER: US/08/049,460
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5646250and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31340
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 713 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-332-638-62

Query Match      15.2%; Score 657.5; DB 1; Length 713;
Best Local Similarity 28.7%; Pred. No. 1.2e-49;
Matches 199; Conservative 104; Mismatches 285; Indels 105; Gaps 23;

Qy 160 NGQYYQIVQIPLMNNVYFQINKTKAISL---TREGSOELNPAKNPSYN-----LVI 211
Db 61 NDKLRYEV-----SSPYFKVNSDGLVALRNITAVGKTLFVHARTPHAEDEMAELVIV 112
Qy 212 SVKDMGQSEN--SFSDDTSV-----DIIVTENIWKAPKPVEMVENSDDPHIKITOVWR 264
Db 113 GKGDIQSLQDIFKARTSPVPROKRSIVS-----PILIPENQRPDRVGKV-- 162
Qy 265 NDPGAQYSLVDKEKLPRFPF-----SIDQE-----GDIYVTPQLDREKDAYVF 308
Db 163 -----VSDRSPERSKFRLTGKGVQDEPKGFRINENTGSVSVTRTLDREVIAYVOL 213
Qy 309 YAVAKDEYKPLSVLEIHKVVKDINDNPPCTPSPVTVFEVQENRGLNSIGTLTAHRRD 368
Db 214 FVETDVGKTLGEGVPLEVIVDQNDNRPIFREGPYIGHVMEGSPGTGTTVMRTAFDAD 273
Qy 369 EENTANSFLNRYVEQTPKLPMDGLFLIQTAYG--MLQAKOSLKKQDT---POYNLTIEV 424
Db 274 DPATDNALLRYNIRQQTDPKPSNMFYIDPEKGDIVTVVSPALLDRETLNPKYELLIEA 333
Qy 425 SDK-----DFKTLCFVQINVIDINDQIPIFEKSDYGNLTJLAEDTNIGSTILTIQATDAE 479
Db 334 QDMAGLDVGLGTATATIMIDDKNDHSPKFTKKEF--QATVEEGAVG--VIVNLTVEDKDD 390
Qy 480 PFTGSSKILYHIKGDSEGRIGVTDPHNTGYVLIKKPLDFETAANSVIVFKAENPEPL 539
Db 391 PTTGAWRAAYTIINGNPGQSFIEHTNPQTNQNEGLSVKPLDYEISAFHTLLIKYENEDPL 450
Qy 540 VFGVKNASSFAKFTLIIVTDVNEAPQFSQHVFOAKVSDVAIGTKVGNVTAKDPEGLD-- 597
Db 451 VPDVSGPSSTATVHITVLDVNEGVPYPPDMVMYTRQEDLSVGSVLLTVNATDPSLQHQ 510
Qy 598 -ISYSLRGDTRGWLKIDHVTGEIFSAPLDREA----GSPYRVQVWATEVGGSSLSVSE 652
Db 511 TIRSVYKDPAGWLNINPINGTVDTTAVLDRESFPVDSVYTFALFLAIDSNPATGTGT 570
Qy 653 FHLIMDVNDNPPRLAKDYTGFLFCHPLSAP-----GSLIFEATDDDOHLFRGPHFT 704
Db 571 LLITLEDVNDNAP-----FIYPTVAEVCDDAKNLSVILGASDKDLHPNTDP--FK 619
Qy 705 FSLGSGSLOND--WEVSKINGTHARLSTRHTDPEERAYVVLIRINDGGRPPLEGIVSLPVT 763

```

```

:|||| : : |||| || ||:: | | : : |||| | | : : |
Db 291 IDFTNRMFLVTAAEANOVPFLAGIOHPPOSTATVSVTVIDVNAENPYFAPNPKIIRQEEG 350

Qy 579 VAIGTKYGNVYAKDPEGL--DISYSLRGDTRGWLKIDHVTGEIFSAPLDREA---GS 631

Db 351 LHAGTMTLTTAQDPDRYMOONTRYTKLSDPANWLKIDPVNGQITTVIAYLDRESPIVQNN 410

Qy 632 PYRVQVATEYGGSSLSVSEFHILMDVNDNPPRLAKDYTGLEFCHPLSA-----PG 684
| :| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 411 IYNATELASDNGIPPMSSGTGLQIYLLDINDNAQVVL-----PQEAETCETPEPN 460

Qy 685 SLIFEATDDOHLRGRPH-TFFSLGSSGLONDNEVSKINGTHARLSTRHTDFFERAYVVL 743
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 461 SINTAALDYDIDPNAGPFAFDPLSPVYTKIRNWTINRLNGDFAQLNLKIFLEAGIYEV 520

Qy 744 IRINDGRRPPLGIVSLPVTFCSC-VBGS 772

Db 521 IIITDSGNPPKSNISILRVKVCQDSMGDC 550

RESULT 13
US-08-474-067-7
; Sequence 7, Application US/08474067
; Patent No. 5811518
; GENERAL INFORMATION:
; APPLICANT: Ranscht, Barbara
; TITLE OF INVENTION: T-Cadherin Adhesion Molecule
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,067
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/213,361
; FILING DATE: 14-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/607,293
; FILING DATE: 30-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1682
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 837 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-474-067-7

Query Match 14.7%; Score 634; DB 2; Length 837;
Best Local Similarity 30.4%; Pred. No. 2e-47;
Matches 189; Conservative 104; Mismatches 261; Indels 68; Gaps

Qy 239 WKAPKPEWENSDTHPIKITYRWNN--DPGAQYSLVDK--EKLPRFPFSDIQE-GDI 292
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 112 WVIP-PISCLNHRGPIYRMRLVQIKSNKDKESKYYVITGGDASPVPVGIFITIEREGWL 170

Qy 293 YVTQPLDREEDKDAYVFYAVAKDEYCKPLSYPLETHVKVKDINDNPPCTCPSPTVFEVQEN 352

```

```

; REFERENCE/DOCKET NUMBER: P-LJ 1683
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 837 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-474-068A-7

Query Match 14.7% Score 634 DB 2 Length 837
Best Local Similarity 30.4% Pred. No. 2e-47
Matches 189; Conservative 104; Mismatches 261; Indels 68; Gaps 25;

QY 239 WKAKPKPVMENVSDPHPIKITQVRWN---DPGAQYSIVDK--EKLPRFPFSIDQE-GDI 292
DB 112 WVIP-PISCLNHRGPPYMLVQIKSNKDKSVYISITGOGADSPVGVIFIERETGWL 170
QY 293 YVTPDLREKDAYVAVAKDEYKGLPSYLEIHKVVKDINDNPPTCPSPVTVFEQEN 352
DB 171 EVTEQLDREKIDRYTLLSHAVSASQPVEDPMEIITVMDQDNKPVFIKEVFVYIEEN 230
QY 353 ERLGNSIGTLTAHDDRDEE-NTANSLFNAVRIVEQTPKLPMDGLFLIQTIVAGMLQAKSLK 411
DB 231 AKPTSVMTVMNATDADAVNTDNGIVSIVSQPPRPHQPFIDPAKGIISVLGTGLD 290
QY 412 KQDTPQYNLTIEVSKDFKTL---CFVQINVIDINDQIPIFEKSDYGNLTIAEDTNGST 468
DB 291 RETPNTVTLVQATDQEGKGLSNTATAIIEVTDANDNIPFNPTMYEG--VVEENKPKTE 348
QY 469 ILTQATDADPEFFGSSKILYHIKGDSEGLVDTDPHTNTGYVIKKPLDFETAASN 528
DB 349 VARITVTDQDAPGSPAWQAVYHIKSGNLDGAFSIITDPSTNNGILKTAKGLDYETKRYD 408
QY 529 IVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFOAKVSEDVAIGTKVGNV 588
DB 409 LVTVENKVPPLSVPI---TLSTASVLVTVLDVNEPPVFPPIKRVGVPEDLPVGOQVTSY 465
QY 589 TAKDPE---GLDISYSLRGDTRGWLKIDHVTGEIPIFSVAPLDREA----GSPYRVQVATE 641
DB 466 TAEDPDRMOKITYRMGSDPAGWLYIHPENGIVTATQPLDRESVHAINSTYKAILAVD 525
QY 642 VGGSSLSVSEFHLIMDVNDNPPRLAKDYTGFLFCFCHPLSAPGSLIFEATDDQHLFRGP 701
DB 526 NGIPDTTGTGTLTLLQLQVNDNGP--TPERSFEICS--RQPEKOILSIVDKD---LPP 577
QY 702 H---FTFSLGSGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGRRPPELEG-- 756
DB 578 HTPFKAALHGS--SNNWTV-EIRGQDELAMGLKKELEPGEYNIFVKLTDS-----QGKA 630
QY 757 -IVSLPVTFCSCVEG---SCFRPA----GHQTGIPTV-GMAVGILLTLLVIGIILAVVF 807
DB 631 QVTQVKAQVCEC-EGTAKNCERRSYIVGG--LGVPAILGILGGIL---ALLILLLLLLLF 684
QY 808 IRIKK-----DKGKDNV 819
DB 685 ARRRKVEKEPLLPEDDMRDNV 706

RESULT 14
US-08-474-068A-7
; Sequence 7, Application US/08474068A
; Patent No. 5837525
; GENERAL INFORMATION:
; APPLICANT: Ranscht, Barbara
; TITLE OF INVENTION: T-Cadherin Adhesion Molecule
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,068A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/213,361
; FILING DATE: 14-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/607,293
; FILING DATE: 30-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
```

```

; REFERENCE/DOCKET NUMBER: P-LJ 1683
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 837 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-474-068A-7

Query Match 14.7% Score 634 DB 2 Length 837
Best Local Similarity 30.4% Pred. No. 2e-47
Matches 189; Conservative 104; Mismatches 261; Indels 68; Gaps 25;

QY 239 WKAKPKPVMENVSDPHPIKITQVRWN---DPGAQYSIVDK--EKLPRFPFSIDQE-GDI 292
DB 112 WVIP-PISCLNHRGPPYMLVQIKSNKDKSVYISITGOGADSPVGVIFIERETGWL 170
QY 293 YVTPDLREKDAYVAVAKDEYKGLPSYLEIHKVVKDINDNPPTCPSPVTVFEQEN 352
DB 171 EVTEQLDREKIDRYTLLSHAVSASQPVEDPMEIITVMDQDNKPVFIKEVFVYIEEN 230
QY 353 ERLGNSIGTLTAHDDRDEE-NTANSLFNAVRIVEQTPKLPMDGLFLIQTIVAGMLQAKSLK 411
DB 231 AKPTSVMTVMNATDADAVNTDNGIVSIVSQPPRPHQPFIDPAKGIISVLGTGLD 290
QY 412 KQDTPQYNLTIEVSKDFKTL---CFVQINVIDINDQIPIFEKSDYGNLTIAEDTNGST 468
DB 291 RETPNTVTLVQATDQEGKGLSNTATAIIEVTDANDNIPFNPTMYEG--VVEENKPKTE 348
QY 469 ILTQATDADPEFFGSSKILYHIKGDSEGLVDTDPHTNTGYVIKKPLDFETAASN 528
DB 349 VARITVTDQDAPGSPAWQAVYHIKSGNLDGAFSIITDPSTNNGILKTAKGLDYETKRYD 408
QY 529 IVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFOAKVSEDVAIGTKVGNV 588
DB 409 LVTVENKVPPLSVPI---TLSTASVLVTVLDVNEPPVFPPIKRVGVPEDLPVGOQVTSY 465
QY 589 TAKDPE---GLDISYSLRGDTRGWLKIDHVTGEIPIFSVAPLDREA----GSPYRVQVATE 641
DB 466 TAEDPDRMOKITYRMGSDPAGWLYIHPENGIVTATQPLDRESVHAINSTYKAILAVD 525
QY 642 VGGSSLSVSEFHLIMDVNDNPPRLAKDYTGFLFCFCHPLSAPGSLIFEATDDQHLFRGP 701
DB 526 NGIPDTTGTGTLTLLQLQVNDNGP--TPERSFEICS--RQPEKOILSIVDKD---LPP 577
QY 702 H---FTFSLGSGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGRRPPELEG-- 756
DB 578 HTPFKAALHGS--SNNWTV-EIRGQDELAMGLKKELEPGEYNIFVKLTDS-----QGKA 630
QY 757 -IVSLPVTFCSCVEG---SCFRPA----GHQTGIPTV-GMAVGILLTLLVIGIILAVVF 807
DB 631 QVTQVKAQVCEC-EGTAKNCERRSYIVGG--LGVPAILGILGGIL---ALLILLLLLLLF 684
QY 808 IRIKK-----DKGKDNV 819
DB 685 ARRRKVEKEPLLPEDDMRDNV 706

RESULT 15
US-08-472-481-6
; Sequence 6, Application US/08472481
; Patent No. 5863804
; GENERAL INFORMATION:
; APPLICANT: Ranscht, Barbara
; TITLE OF INVENTION: T-Cadherin Adhesion Molecule
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
```



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 2, 2003, 15:54:14 ; Search time 42.7397 Seconds  
(without alignments)  
1190.115 Million cell updates/sec

Title: US-09-079-678-178  
Perfect score: 4321  
Sequence: 1 MIIQAHLSCLLMYLATG.....DKGRDNVESQAQSEVKPLRS 832

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 248812 seqs, 61136040 residues

Total number of hits satisfying chosen parameters: 248812

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : Published Applications\_AA.\*
- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pap.\*
  - 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pap.\*
  - 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pap.\*
  - 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pap.\*
  - 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pap.\*
  - 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pap.\*
  - 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pap.\*
  - 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pap.\*
  - 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pap.\*
  - 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pap.\*
  - 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pap.\*
  - 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pap.\*
  - 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pap.\*
  - 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	4321	100.0	832	9	US-10-025-380-1081
2	4321	100.0	832	10	US-09-922-217-1081
3	4321	100.0	832	10	US-09-833-263-1081
4	4208	97.4	958	9	US-10-025-380-1087
5	4208	97.4	958	10	US-09-922-217-1087
6	4208	97.4	958	10	US-09-833-263-1087
7	908.5	21.0	807	9	US-10-063-547-98
8	908.5	21.0	807	9	US-10-063-616-98
9	908.5	21.0	807	9	US-10-063-502-98
10	908.5	21.0	807	9	US-10-227-884-188
11	908.5	21.0	807	9	US-10-230-163-188
12	908.5	21.0	807	9	US-10-006-856A-229
13	908.5	21.0	807	9	US-10-218-631-188
14	908.5	21.0	807	9	US-10-230-338-188
15	908.5	21.0	807	9	US-10-063-518-98
16	908.5	21.0	807	9	US-10-063-598-98
17	908.5	21.0	807	9	US-10-227-693-98
18	908.5	21.0	807	9	US-10-230-414-188
19	908.5	21.0	807	9	US-10-006-818A-229

ALIGNMENTS

RESULT 1

US-10-025-380-1081  
; Sequence 1081, Application US/10025380  
; Publication No. US20020182191A1  
; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Meagher, Madeleine Joy  
; APPLICANT: Stolk, John A.  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Jiang, Yuqiu  
; APPLICANT: Smith, Carole L.  
; APPLICANT: King, Gordon E.  
; APPLICANT: Wang, Aijun  
; APPLICANT: Clapper, Jonathan D.  
; APPLICANT: Skeiky, Yasir A. W.  
; APPLICANT: Fanger, Gary R.  
; APPLICANT: Vedvick Thomas S.  
; APPLICANT: Carter, Darrick  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS  
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE  
; FILE REFERENCE: 210121.471C14  
; CURRENT APPLICATION NUMBER: US/10/025,380  
; CURRENT FILING DATE: 2001-12-19  
; NUMBER OF SEQ ID NOS: 1129  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1081  
; LENGTH: 832  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-025-380-1081

Query Match 100.0%; Score 4321; DB 9; Length 832;  
Best Local Similarity 100.0%; Pred. No. 2e-314;  
Matches 832; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MIIQAHLSCLLMYLATGYGQEGKFSGLPKMPTFSIYEGQEPSQLIFQKAPPAVTF 60  
Db 1 MIIQAHLSCLLMYLATGYGQEGKFSGLPKMPTFSIYEGQEPSQLIFQKAPPAVTF 60

Sequence 98, Appl  
Sequence 5, Appl  
Sequence 46, Appl  
Sequence 2, Appl  
Sequence 46, Appl  
Sequence 49, Appl  
Sequence 48, Appl  
Sequence 47, Appl  
Sequence 13, Appl  
Sequence 14, Appl  
Sequence 5, Appl  
Sequence 2, Appl  
Sequence 3, Appl  
Sequence 48, Appl  
Sequence 2, Appl  
Sequence 25, Appl  
Sequence 30, Appl  
Sequence 6, Appl  
Sequence 6, Appl  
Sequence 29, Appl  
Sequence 21, Appl  
Sequence 11, Appl  
Sequence 237, Appl  
Sequence 2, Appl  
Sequence 55, Appl

```
Qy 61 ELTGTDNIFVIERGLLYNRRALDRSTRTHNLQVAALDANGIIVEGVPVITIEVKDIN 120
Db 61 ELTGTDNIFVIERGLLYNRRALDRSTRTHNLQVAALDANGIIVEGVPVITIEVKDIN 120
Qy 121 DNRPTFLQSKYEGSVQRNSRPGKPFLLYNATDLDPATPNQGLYYQIVQLPMNNMYF 180
Db 121 DNRPTFLQSKYEGSVQRNSRPGKPFLLYNATDLDPATPNQGLYYQIVQLPMNNMYF 180
Qy 181 QINNKGTGAISLTRGSOELNPAKNSYNLVISVKDMGGQSENSESDTTSVDIIIVTENLWK 240
Db 181 QINNKGTGAISLTRGSOELNPAKNSYNLVISVKDMGGQSENSESDTTSVDIIIVTENLWK 240
Qy 241 APKPVEMVENSTDPHPKIKITQVRWNDPCAQYSLVDKKEKLPDPFETAAVSNIVFKAENPEPLV 300
Db 241 APKPVEMVENSTDPHPKIKITQVRWNDPCAQYSLVDKKEKLPDPFETAAVSNIVFKAENPEPLV 300
Qy 301 EKDAYVFYAVAKDEYKPLSYPLEIHVKVKDINDNPPCTCSPVTVFEVQENERLGNISG 360
Db 301 EKDAYVFYAVAKDEYKPLSYPLEIHVKVKDINDNPPCTCSPVTVFEVQENERLGNISG 360
Qy 361 TLTAHRRDEENTANSFLNRYVEQTPKLPMDGLFLIOTYAGMLQAKOSLKKQDTPQYNL 420
Db 361 TLTAHRRDEENTANSFLNRYVEQTPKLPMDGLFLIOTYAGMLQAKOSLKKQDTPQYNL 420
Qy 421 TIEVSDKDFKTLCFVQINVIDNQIPIFEKSDYGNLTAEEDTNGSTILTIQATDADEP 480
Db 421 TIEVSDKDFKTLCFVQINVIDNQIPIFEKSDYGNLTAEEDTNGSTILTIQATDADEP 480
Qy 481 FTGSSKILYHIKGDSEGRGCVDPDHTNTGYVLIKKPLDFETAAVSNIVFKAENPEPLV 540
Db 481 FTGSSKILYHIKGDSEGRGCVDPDHTNTGYVLIKKPLDFETAAVSNIVFKAENPEPLV 540
Qy 541 FGVYNASSFAKFTLIIVTDVNEAPQFSOHVFOAKVSEDVAIGTKVGNVTAKDPEGLDISY 600
Db 541 FGVYNASSFAKFTLIIVTDVNEAPQFSOHVFOAKVSEDVAIGTKVGNVTAKDPEGLDISY 600
Qy 601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVADEVGSSLSVSEFHLILMDV 660
Db 601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVADEVGSSLSVSEFHLILMDV 660
Qy 661 NDNPRLAKDVTGLFFCHPLSAPGLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEYSK 720
Db 661 NDNPRLAKDVTGLFFCHPLSAPGLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEYSK 720
Qy 721 INGTARLSTRHTDFEERAYVVLIRINDGRRPPLLEGIVSLPVTFCSCVEGSCFRPAGHOT 780
Db 721 INGTARLSTRHTDFEERAYVVLIRINDGRRPPLLEGIVSLPVTFCSCVEGSCFRPAGHOT 780
Qy 781 GIPTVGMVAGILLTLLVIGIILAVFIRIKKDKGKNVESQAQSEVKPLRS 832
Db 781 GIPTVGMVAGILLTLLVIGIILAVFIRIKKDKGKNVESQAQSEVKPLRS 832
```

## RESULT 2

```
US-09-922-217-1081
; Sequence 1081, Application US/09922217
; Patent No. US2002007641A1
```

## GENERAL INFORMATION:

```
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Jiang, Yugu
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C13
```

```
RESULT 3
US-09-833-263-1081
```

```
; CURRENT APPLICATION NUMBER: US/09/922.217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1081
; LENGTH: 832
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-217-1081
```

```
Query Match 100.0%; Score 4321; DB 10; Length 832;
Best Local Similarity 100.0%; Pred. No. 2e-314;
Matches 832; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 MILQAHLSLCLMLLYLATGYGQSGKPSGLKPMPTFSIYEGQEPSQIIIFQKANPPAVTF 60
Db 1 MILQAHLSLCLMLLYLATGYGQSGKPSGLKPMPTFSIYEGQEPSQIIIFQKANPPAVTF 60
Qy 61 ELTGTONIFVIERGLLYNRRALDRSTRTHNLQVAALDANGIIVEGVPVITIEVKDIN 120
Db 61 ELTGTONIFVIERGLLYNRRALDRSTRTHNLQVAALDANGIIVEGVPVITIEVKDIN 120
Qy 121 DNRPTFLQSKYEGSVQRNSRPGKPFLLYNATDLDPATPNQGLYYQIVQLPMNNMYF 180
Db 121 DNRPTFLQSKYEGSVQRNSRPGKPFLLYNATDLDPATPNQGLYYQIVQLPMNNMYF 180
Qy 181 QINNKGTGAISLTRGSOELNPAKNSYNLVISVKDMGGQSENSESDTTSVDIIIVTENLWK 240
Db 181 QINNKGTGAISLTRGSOELNPAKNSYNLVISVKDMGGQSENSESDTTSVDIIIVTENLWK 240
Qy 241 APKPVEMVENSTDPHPKIKITQVRWNDPCAQYSLVDKKEKLPDPFETAAVSNIVFKAENPEPLV 300
Db 241 APKPVEMVENSTDPHPKIKITQVRWNDPCAQYSLVDKKEKLPDPFETAAVSNIVFKAENPEPLV 300
Qy 301 EKDAYVFYAVAKDEYKPLSYPLEIHVKVKDINDNPPCTCSPVTVFEVQENERLGNISG 360
Db 301 EKDAYVFYAVAKDEYKPLSYPLEIHVKVKDINDNPPCTCSPVTVFEVQENERLGNISG 360
Qy 361 TLTAHRRDEENTANSFLNRYVEQTPKLPMDGLFLIOTYAGMLQAKOSLKKQDTPQYNL 420
Db 361 TLTAHRRDEENTANSFLNRYVEQTPKLPMDGLFLIOTYAGMLQAKOSLKKQDTPQYNL 420
Qy 421 TIEVSDKDFKTLCFVQINVIDNQIPIFEKSDYGNLTAEEDTNGSTILTIQATDADEP 480
Db 421 TIEVSDKDFKTLCFVQINVIDNQIPIFEKSDYGNLTAEEDTNGSTILTIQATDADEP 480
Qy 481 FTGSSKILYHIKGDSEGRGCVDPDHTNTGYVLIKKPLDFETAAVSNIVFKAENPEPLV 540
Db 481 FTGSSKILYHIKGDSEGRGCVDPDHTNTGYVLIKKPLDFETAAVSNIVFKAENPEPLV 540
Qy 541 FGVYNASSFAKFTLIIVTDVNEAPQFSOHVFOAKVSEDVAIGTKVGNVTAKDPEGLDISY 600
Db 541 FGVYNASSFAKFTLIIVTDVNEAPQFSOHVFOAKVSEDVAIGTKVGNVTAKDPEGLDISY 600
Qy 601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVADEVGSSLSVSEFHLILMDV 660
Db 601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVADEVGSSLSVSEFHLILMDV 660
Qy 661 NDNPRLAKDVTGLFFCHPLSAPGLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEYSK 720
Db 661 NDNPRLAKDVTGLFFCHPLSAPGLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEYSK 720
Qy 721 INGTARLSTRHTDFEERAYVVLIRINDGRRPPLLEGIVSLPVTFCSCVEGSCFRPAGHOT 780
Db 721 INGTARLSTRHTDFEERAYVVLIRINDGRRPPLLEGIVSLPVTFCSCVEGSCFRPAGHOT 780
Qy 781 GIPTVGMVAGILLTLLVIGIILAVFIRIKKDKGKNVESQAQSEVKPLRS 832
Db 781 GIPTVGMVAGILLTLLVIGIILAVFIRIKKDKGKNVESQAQSEVKPLRS 832
```

```
; Sequence 1081, Application US/09833263
; Patent No. US20020110547A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Stolk, John A.
; APPLICANT: Meagher, Madeleine J.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C12
; CURRENT APPLICATION NUMBER: US/09/833,263
; CURRENT FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 1093
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1081
; LENGTH: 832
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-263-1081

Query Match      100.0%; Score 4321; DB 10; Length 832;
Best Local Similarity 100.0%; Pred. No. 2e-314;
Matches 832; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MILQAHLSLCLMLYLATYGGQEGKFSGLPKMTFSIYEGQBPSSQIIFOFKANPPAVTF 60
DB 1 MILQAHLSLCLMLYLATYGGQEGKFSGLPKMTFSIYEGQBPSSQIIFOFKANPPAVTF 60
QY 61 ELTGETDNFVIERGLLYNRALDRETRSTHNLQVAALDANGIIVGPPVITIEVKDIN 120
DB 61 ELTGETDNFVIERGLLYNRALDRETRSTHNLQVAALDANGIIVGPPVITIEVKDIN 120
QY 121 DNRPTFLQSKYEGSVQRNSRPGKPFYLVNATDLDPATNGQLYYQIVIQLPMMNVMYF 180
DB 121 DNRPTFLQSKYEGSVQRNSRPGKPFYLVNATDLDPATNGQLYYQIVIQLPMMNVMYF 180
QY 181 QINNKTAISLREGSOELNPAKNPSYLVSKMGQSGSEFSFDTTSVDIIIVTENIWK 240
DB 181 QINNKTAISLREGSOELNPAKNPSYLVSKMGQSGSEFSFDTTSVDIIIVTENIWK 240
QY 241 AKPVEWENSTDPHPKIKITQVRWNDPGAQYSLVDKEKLPFRPFSDIQEGDIYVTOPLDR 300
DB 241 AKPVEWENSTDPHPKIKITQVRWNDPGAQYSLVDKEKLPFRPFSDIQEGDIYVTOPLDR 300
QY 301 EKDQAVFYAVAKDEYKPLSYPLEIHVKVDINDNPPCTCSPVTVFEQENERLGNISG 360
DB 301 EKDQAVFYAVAKDEYKPLSYPLEIHVKVDINDNPPCTCSPVTVFEQENERLGNISG 360
QY 361 TLTAHREDEANTANSEFNRIEVTQPKLPMDFGLFIQTYAGMLQAKSLKQDTPQYNL 420
DB 361 TLTAHREDEANTANSEFNRIEVTQPKLPMDFGLFIQTYAGMLQAKSLKQDTPQYNL 420
QY 421 TIEVSDKDKFTLCFQVINIDINDQIPIEKSQYGNLTAEEDNIGSTILTQATDADEP 480
DB 421 TIEVSDKDKFTLCFQVINIDINDQIPIEKSQYGNLTAEEDNIGSTILTQATDADEP 480
QY 481 FTGSSKILYHIKGDSEGRGLVDTPHTNTGYVYIIKKPLDFFETAASNTIVFKAENPEPLV 540
DB 481 FTGSSKILYHIKGDSEGRGLVDTPHTNTGYVYIIKKPLDFFETAASNTIVFKAENPEPLV 540
QY 541 FGKYYNASSFAKFTLIVTDVNEAPQSQHVFOAKVSEDAIGTKVGNVTAKDPEGLDISY 600
DB 541 FGKYYNASSFAKFTLIVTDVNEAPQSQHVFOAKVSEDAIGTKVGNVTAKDPEGLDISY 600
QY 601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRQOVVATEVGGSSLSVSEPHLLIMDV 660
DB 601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRQOVVATEVGGSSLSVSEPHLLIMDV 660
QY 661 NDNPPLAKDYTGGLFCHPLSAPGSLFEATDDQHLFRGPHFTFSLGSGSLQNDNEVSK 720
DB 661 NDNPPLAKDYTGGLFCHPLSAPGSLFEATDDQHLFRGPHFTFSLGSGSLQNDNEVSK 720
QY 721 INGTHARLSTRHTDFEERAYVVLIRINDGGRPLEGIVSLPVTFCSCVGCSPFRPAGHOT 780
```

```
DB 721 INGTHARLSTRHTDFEERAYVVLIRINDGGRPLEGIVSLPVTFCSCVGCSPFRPAGHOT 780
QY 781 GIPTVGMVAGILLTLLVIGIILAVVFIRIKDKGKDNVESQAQASEVKPLRS 832
DB 781 GIPTVGMVAGILLTLLVIGIILAVVFIRIKDKGKDNVESQAQASEVKPLRS 832

RESULT 4
US-10-025-380-1087
; Sequence 1087, Application US/10025380
; Publication No. US20020182191A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Smith, Carole L.
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darick
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C14
; CURRENT APPLICATION NUMBER: US/10/025,380
; CURRENT FILING DATE: 2001-12-19
; NUMBER OF SEQ ID NOS: 1129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1087
; LENGTH: 958
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-025-380-1087

Query Match      97.4%; Score 4208; DB 9; Length 958;
Best Local Similarity 99.4%; Pred. No. 6.7e-306;
Matches 813; Conservative 1; Mismatches 2; Indels 2; Gaps 1;

QY 17 LATG--YGOEGKFSGLPKMTFSIYEGQBPSSQIIFOFKANPPAVTFELTGETDNIFVIER 74
DB 141 LASGSAAGKFSGLPKMTFSIYEGQBPSSQIIFOFKANPPAVTFELTGETDNIFVIER 200
QY 75 EGLLYNRALDRETRSTHNLQVAALDANGIIVGPPVITIEVKDINDNRPFTFLQSKYEGS 134
DB 201 EGLLYNRALDRETRSTHNLQVAALDANGIIVGPPVITIEVKDINDNRPFTFLQSKYEGS 260
QY 135 VRONSRRPGKPFYLVNATDLDPATNGQLYYQIVIQLPMMINNYMFIQNKTKAISLTRE 194
DB 261 VRONSRRPGKPFYLVNATDLDPATNGQLYYQIVIQLPMMINNYMFIQNKTKAISLTRE 320
QY 195 GSQELNPAKNPSYLVSKMGQSGSEFSFDTTSVDIIIVTENIWKAPKPEVWENSTDP 254
DB 321 GSQELNPAKNPSYLVSKMGQSGSEFSFDTTSVDIIIVTENIWKAPKPEVWENSTDP 380
QY 255 HPKIKITQVRWNDPGAQYSLVDKEKLPFRPFSDIQEGDIYVTOPLDREKDAYFYAVAKD 314
DB 381 HPKIKITQVRWNDPGAQYSLVDKEKLPFRPFSDIQEGDIYVTOPLDREKDAYFYAVAKD 440
QY 315 EYKGPLSYPLEIHVKVDINDNPPCTCSPVTVFEQENERLGNISIGTLTAHREDEANTAN 374
DB 441 EYKGPLSYPLEIHVKVDINDNPPCTCSPVTVFEQENERLGNISIGTLTAHREDEANTAN 500
QY 375 SFUNYRIVEQTPKLPMDGFLFIQTYAGMLQAKSLKQDTPQYNLTIEVSKDKFTLCF 434
DB 501 SFUNYRIVEQTPKLPMDGFLFIQTYAGMLQAKSLKQDTPQYNLTIEVSKDKFTLCF 560
```

QY 435 VQINVIDINDQIPIFEKSDYGNLTAEADTNGISTILTIQATDADEPFTGSSKILYHIIG 494  
Db 561 VQINVIDINDQIPIFEKSDYGNLTAEADTNGISTILTIQATDADEPFTGSSKILYHIIG 620  
QY 495 DSEGRGVDPHTNTGYVLIKKPLDPFETAAVSNIVFKAENPELVFGVKYNASSFAKFT 554  
Db 621 DSEGRGVDPHTNTGYVLIKKPLDPFETAAVSNIVFKAENPELVFGVKYNASSFAKFT 680  
QY 555 LIVTDVNEAPQFSOHVFOAKVSEDVAIGTKVGNVTAKDPEGLDISYSLRGDTRGWLKIDH 614  
Db 681 LIVTDVNEAPQFSOHVFOAKVSEDVAIGTKVGNVTAKDPEGLDISYSLRGDTRGWLKIDH 740  
QY 615 VTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSVSEFHLILMDVNDNPPRLAKDYTGL 674  
Db 741 VTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSVSEFHLILMDVNDNPPRLAKDYTGL 800  
QY 675 FFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEVSKINGTHARLSTRHTD 734  
Db 801 FFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEVSKINGTHARLSTRHTD 860  
QY 735 FEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQTGIPVGMVAVGILLT 794  
Db 861 FEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQTGIPVGMVAVGILLT 920  
QY 795 TLLVIGIILAVFIRIKKDKGKDNVESQAQSEVKPLRS 832  
Db 921 TLLVIGIILAVFIRIKKDKGKDNVESQAQSEVKPLRS 958

RESULT 5

US-09-922-217-1087  
; Sequence 1087, Application US/09922217  
; Patent No. US20020076414A1  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Meagher, Madeleine Joy  
; APPLICANT: Stolk, John A.  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Jiang, Yuqiu  
; APPLICANT: Smith, Carole Lynn  
; APPLICANT: King, Gordon E.  
; APPLICANT: Wang, Aijun  
; APPLICANT: Clapper, Jonathan D.  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS  
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE  
; FILE REFERENCE: 210121.471C13  
; CURRENT APPLICATION NUMBER: US/09/922,217  
; CURRENT FILING DATE: 2001-08-03  
; NUMBER OF SEQ ID NOS: 1124  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1087  
; LENGTH: 958  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-922-217-1087

Query Match 97.4%; Score 4208; DB 10; Length 958;  
Best Local Similarity 99.4%; Pred. No. 6.7e-306;  
Matches 813; Conservative 1; Mismatches 2; Indels 2; Gaps 1;

QY 17 LATG--YGOEGKSGPLKPMFTSEIYEGQPSQIIFQKANNPPAVTFELTGETDNIFVIER 74  
Db 141 LASGAAQAGKSGPLKPMFTSEIYEGQPSQIIFQKANNPPAVTFELTGETDNIFVIER 200  
QY 75 EGLLYNRALDRSTRSHNLQVAALDANGIIVEGVPITIEVKDINDNRPTFLOSKYEGS 134  
Db 201 EGLLYNRALDRSTRSHNLQVAALDANGIIVEGVPITIEVKDINDNRPTFLOSKYEGS 260  
QY 135 VRQNSRPGKPFLLVYNATDLDDPATPNQQLYYQIVQLPMMNNVYFOINNKTGAISLTRE 194

Db 261 VRQNSRPGKPFLLVYNATDLDDPATPNQQLYYQIVQLPMMNNVYFOINNKTGAISLTRE 320  
QY 195 GSQELNPAKNPSNLYISVKDMGGQSENSESDTTSVDIIIVTENIWKAPKPEVVENSTDP 254  
Db 321 GSQELNPAKNPSNLYISVKDMGGQSENSESDTTSVDIIIVTENIWKAPKPEVVENSTDP 380  
QY 255 HPKIKITQVRWNDPGAQYSLVDKEKLPFRPFPSIDQEGDIYVYVQPLDREKDAYVYFAVAKD 314  
Db 381 HPKIKITQVRWNDPGAQYSLVDKEKLPFRPFPSIDQEGDIYVYVQPLDREKDAYVYFAVAKD 440  
QY 315 EYCKPLSYPLEIIVHKVDINDNPCTPSPVTVEQENERLGNISIGTLTAHDDEENTAN 374  
Db 441 EYCKPLSYPLEIIVHKVDINDNPCTPSPVTVEQENERLGNISIGTLTAHDDEENTAN 500  
QY 375 SFNLRYIVEQTPKLPMDGLFIQIYAGMLQAKQSLKKQDTPQYNLTIEVSDKDFKTLCF 434  
Db 501 SFNLRYIVEQTPKLPMDGLFIQIYAGMLQAKQSLKKQDTPQYNLTIEVSDKDFKTLCF 560  
QY 435 VQINVIDINDQIPIFEKSDYGNLTAEADTNGISTILTIQATDADEPFTGSSKILYHIIG 494  
Db 561 VQINVIDINDQIPIFEKSDYGNLTAEADTNGISTILTIQATDADEPFTGSSKILYHIIG 620  
QY 495 DSEGRGVDPHTNTGYVLIKKPLDPFETAAVSNIVFKAENPELVFGVKYNASSFAKFT 554  
Db 621 DSEGRGVDPHTNTGYVLIKKPLDPFETAAVSNIVFKAENPELVFGVKYNASSFAKFT 680  
QY 555 LIVTDVNEAPQFSOHVFOAKVSEDVAIGTKVGNVTAKDPEGLDISYSLRGDTRGWLKIDH 614  
Db 681 LIVTDVNEAPQFSOHVFOAKVSEDVAIGTKVGNVTAKDPEGLDISYSLRGDTRGWLKIDH 740  
QY 615 VTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSVSEFHLILMDVNDNPPRLAKDYTGL 674  
Db 741 VTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSVSEFHLILMDVNDNPPRLAKDYTGL 800  
QY 675 FFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEVSKINGTHARLSTRHTD 734  
Db 801 FFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEVSKINGTHARLSTRHTD 860  
QY 735 FEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQTGIPVGMVAVGILLT 794  
Db 861 FEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQTGIPVGMVAVGILLT 920  
QY 795 TLLVIGIILAVFIRIKKDKGKDNVESQAQSEVKPLRS 832  
Db 921 TLLVIGIILAVFIRIKKDKGKDNVESQAQSEVKPLRS 958

RESULT 6

US-09-833-263-1087  
; Sequence 1087, Application US/09833263  
; Patent No. US20020110547A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Aijun  
; APPLICANT: Clapper, Jonathan D.  
; APPLICANT: Stolk, John A.  
; APPLICANT: Meagher, Madeleine J.  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND  
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE  
; FILE REFERENCE: 210121.471C12  
; CURRENT APPLICATION NUMBER: US/09/833,263  
; CURRENT FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 1093  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1087  
; LENGTH: 958  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-833-263-1087

Query Match 97.4%; Score 4208; DB 10; Length 958;  
Best Local Similarity 99.4%; Pred. No. 6.7e-306;  
Matches 813; Conservative 1; Mismatches 2; Indels 2; Gaps 1;

Qy 17 LATG--YQGEKFGSPLKPMFTSIYEGQEPSQIIIFQFKANPPAVTFTGTDMIFVIER 74  
Dy 141 LASGSAAEKFGSPLKPMFTSIYEGQEPSQIIIFQFKANPPAVTFTGTDMIFVIER 200  
Qy 75 EGLLYNRRALDRETRSHNLQVAALDANGIIIEGVPVITIEVKDINDNRPTFLQSKYEGS 134  
Dy 201 EGLLYNRRALDRETRSHNLQVAALDANGIIIEGVPVITIEVKDINDNRPTFLQSKYEGS 260  
Qy 135 VRQNSRPGKPLYNATDLDPATPNQOLYYQIYQLPMINNMVYFOINNKGTGALSITRE 194  
Dy 261 VRQNSRPGKPLYNATDLDPATPNQOLYYQIYQLPMINNMVYFOINNKGTGALSITRE 320  
Qy 195 GSQELNPAKNPSYNLVTSVKDMGQSENSEFSDTTSVDIIVTENIWKAPKPVEMVENSTDP 254  
Dy 321 GSQELNPAKNPSYNLVTSVKDMGQSENSEFSDTTSVDIIVTENIWKAPKPVEMVENSTDP 380  
Qy 255 HPIKITQVRWMDPCAQYSLVDKELPRPFPFSDIDEGDIYVTPQDLREKDAYFYAVAKD 314  
Dy 381 HPIKITQVRWMDPCAQYSLVDKELPRPFPFSDIDEGDIYVTPQDLREKDAYFYAVAKD 440  
Qy 315 EYGPVLSYPLIHHVKVDINDNPTPCSPVTVFEQENRGLNSIGTLTAHDEEENTAN 374  
Dy 441 EYGPVLSYPLIHHVKVDINDNPTPCSPVTVFEQENRGLNSIGTLTAHDEEENTAN 500  
Qy 375 SFLNRYRIVEQTPKLPMDGLFLIOTYAGMLQAKOSLKKQDTPQYNLTIEVSDKDFKTLCF 434  
Dy 501 SFLNRYRIVEQTPKLPMDGLFLIOTYAGMLQAKOSLKKQDTPQYNLTIEVSDKDFKTLCF 560  
Qy 435 VOINVIDINDOIPFEKSDYGNLTLAEDNTGISTILTIQATDADEPFTGSSKILYHIK 494  
Dy 561 VOINVIDINDOIPFEKSDYGNLTLAEDNTGISTILTIQATDADEPFTGSSKILYHIK 620  
Qy 495 DSEGRGVDTDPHNTGTVIILKPLDPETAASNTVKAENPEPLVGVKYNASFEAKFT 554  
Dy 621 DSEGRGVDTDPHNTGTVIILKPLDPETAASNTVKAENPEPLVGVKYNASFEAKFT 680  
Qy 555 LIVTDVNEAPOFSQHVFOAKYSEDAVIGTKVGNVTAKDPEGLDISYSLRGDTRGWLKIDH 614  
Dy 681 LIVTDVNEAPOFSQHVFOAKYSEDAVIGTKVGNVTAKDPEGLDISYSLRGDTRGWLKIDH 740  
Qy 615 VTGEIFSVAPLDRAGSPYRVQVYVATEVGGSSLSVSEFHLILMDVNDNPPRLAKDYTL 674  
Dy 741 VTGEIFSVAPLDRAGSPYRVQVYVATEVGGSSLSVSEFHLILMDVNDNPPRLAKDYTL 800  
Qy 675 FFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSSLONDWEYSKINGTHARLSTRHTD 734  
Dy 801 FFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSSLONDWEYSKINGTHARLSTRHTD 860  
Qy 735 FEERAYVVLIRINDGRRPPLLEGIVSLPVTFCSCVGGSCFRPAGHOTGIPTVGMVAVGILLT 794  
Dy 861 FEERAYVVLIRINDGRRPPLLEGIVSLPVTFCSCVGGSCFRPAGHOTGIPTVGMVAVGILLT 920  
Qy 795 TLLVIGIILAVVFIIRKKDKNDVESAQAQSEVPLRS 832  
Dy 921 TLLVIGIILAVVFIIRKKDKNDVESAQAQSEVPLRS 958

RESULT 7  
US-10-063-547-98

; Sequence 98, Application US/10063547  
; Publication No. US20020182638A1  
; GENERAL INFORMATION:  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
; FILE REFERENCE: P3230R1C1  
; CURRENT APPLICATION NUMBER: US/10/063,547  
; CURRENT FILING DATE: 2002-05-02  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 170  
; SEQ ID NO 98  
; LENGTH: 807  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-063-547-98

Query Match 21.0%; Score 908.5; DB 9; Length 807;  
Best Local Similarity 30.2%; Pred. No. 1.6e-59;  
Matches 237; Conservative 159; Mismatches 335; Indels 53; Gaps 21;

Qy 62 LTGE-----TDMIFVIERE-GLLYNRRALDRETRSHNLQVAALDANGIIIEGVPVITIEV 116  
Dy 57 LSGDSGKATEGPFAMDPDSGFLVTRALDREEQAEYQVTLQEMQDGHVLMGQPQVPLVHV 116  
Qy 117 KDINDNRPTFLQSKYEGSVRONSRPGKPLVYNATDLDPATPNQOLYYQIYQLPMINN 176  
Dy 117 KDENDQVPHFSQAIYRARLSRGTRGIPFLFLEASDRDEPGTANSLDFHILSQAPQPS 176  
Qy 177 VMYFQINNKTGAISLTREGSQELNPAKNPSYNLVTSVKDMGQSENSEFSDTTSVDIIVTE 236  
Dy 177 PDMFQLEPRLCALALSPKSGTSLDHALERTYQLLVQVQKMDGQA-SGHQATATVEVSIIE 235  
Qy 237 NIWKAPKPVEMVENSTDPHPIKITQVRWMDPCAQYSLVDKELPRPFPFSDIDEGDIYVTO 296  
Dy 236 STWVSLEPIHLAENLKVLYPHHMAQVHWGGDVHYHL---ESHPPGPEVNAEGLNYVTR 292  
Qy 297 PLDREKDAYFYAVAKDEVCKPLSYPLEIHHVKVDINDNPTPCSPVTVFEQENRGL 356  
Dy 293 ELDRQAQAEYLQVRAQNSHGEDIAAPLELHVLVNDENINVCPPRDPPTVSIPELSPG 352  
Qy 357 NSIGTLTAHDEEENTANSLNRYRIVEQTPKLPMDG-LFLIOTYAGMLQAKOSLKK-QD 414  
Dy 353 TEVTRLSAEDADAPGSPNSHVYQLLSPEDGEGVAGRAFOVDPTSGSVTLGLVPLRAQON 412  
Qy 415 TPQYNLTIEVSDKD--FKTLCFVQINVIDINDOIPFEKSDYGNLTLAEDNTGISTILTI 472  
Dy 413 ILLVLVLAIDLAGAGGGFSSCTCEVAVATDINDHAPEFTISQIGPISLPEDVEPGTLVAML 472  
Qy 473 QATDAD-EPFTGSSKIL-YHIIKGDSEGRGVDTDPHNTGTVIILKPLDPETAASNTVKAEN 528  
Dy 473 TADADLEP---APFLMDFAIERGDTGTFGLDWEF--DSGHVRLRCKNLSYEAAPSHE 527  
Qy 529 IVFKAENPEPLVGVKYNASFEAKFTLIVTDVNEAPOFSQHVFOAKYSEDAVIGTKVGNV 588  
Dy 528 VVVVVQSVAKLV-GPGPGGATATVTLVERVMPKLDQESYEASVPISAPAGSFLTI 586  
Qy 589 TAKDPEGLDISYSLRGDTRGWLKIDHVTGEIFSVAPLD-REAGSPYRVQVYVATEVGGSSL 647  
Dy 587 QPSDPISTLRFSLVNDSEGLCEKFSGEVHTAQSLOGAQPGDTYTVLVEAQDTA---- 642  
Qy 648 SSVSEFHLILMDVNDNPPRLAKDYTLFFCHPLSAPGSLIF-EATDDQHLFRGPHFTFS 706  
Dy 643 -----LTLAPVPSQ-----YLCPTRODHGLIVSGSPKDPDLASGHGP-YSET 683  
Qy 707 LG-SGSLQNDWEYSKINGTHARLSTRHTDFFERAYVVLIRINDGRRPPLLEGIVSLPVTFC 765  
Dy 684 LGPNPTVQRDRLQTLNGSHAYLTALHWVEPREHIIPVVVSHNAQ---MQQLLVRVIVC 740  
Qy 766 SC-VEGSCFRPAGHOTGIPTVGMVAVGILLTLLVIGIILAVVFIIRKKDKNDVESAQA 824  
Dy 741 RCNVEGQCMRKVGRMKGMPTKLSAVGILVGLTVAIGIFLILIFTHWTMSRKKDQDPADS 800  
Qy 825 SEVK 828  
Dy 801 VPLK 804

## RESULT 8

```
US-10-063-616-98
; Sequence 98, Application US/10063616
; Publication No. US2003001385A1
; GENERAL INFORMATION:
; APPLICANT: Eaton, Dan L.
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3230R1C1
; CURRENT APPLICATION NUMBER: US/10/063,616
; CURRENT FILING DATE: 2002-05-03
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 98
; LENGTH: 807
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-063-616-98

Query Match      21.0%; Score 908.5; DB 9; Length 807;
Best Local Similarity 30.2%; Pred. No. 1.6e-59;
Matches 237; Conservative 159; Mismatches 335; Indels 53; Gaps 21;

Qy 62 LTGE----TDNIFVIERE-GLLYYNRALDRETRSTHNLQVAALDANGIIIEGPPVPIIEV 116
Db 57 LSGDSGRKATEGPFAMDPDGSGLLVTRALDREEQAEYQLQVTFLEMQDGHVLMGPQVPLVHV 116
Qy 117 KDINDNRPFTLOSKEYSVRONSRPGPFLLVYNATDLDPATNGQLYQIVIQPLMINN 176
Db 117 KDENDQVPHFSQAIYRARLSRGTRPGIPFLFLEASDRDEPTANSDLRFLHLSQAQPS 176
Qy 177 VMYFOINNKTKAISLTREGSQELNPAKNSYNLVISKMGQSGSENSFSDTTSVDIIYTE 236
Db 177 PDMFOLEPRLGALALS PKGSTSLDHALERTYQLLVQVKMDGQA--SGHQATATVEVSII 235
Qy 237 NIWKAPRPVENVSTDPHPKIKITQVRWNPFGAQYSLVDKEKLPFRFFPSIDQEGDIYVTQ 296
Db 236 STWVSLEPIHLAENLKVLVPHHMAQVHWSGGDVHYHL---ESHPPGPFVNAEGLNYVTR 292
Qy 297 PLDREEDKDAYVYAVAKDEYCKPLSYPLEIHVKVDINDNPPTCPSPVTVEVOENERLG 356
Db 293 ELDRQAQAEYLLQVRAQNSHGDEYAAPLEHLVYMDENDNVPICPPRDPVTYSIPELSPG 352
Qy 357 NSIGTLTAHDEENTANSFLNYRIVEQTPKLPMDG-LFLIQTVAGMLQAKOSLKK-OD 414
Db 353 TEVTRLSAEDADAPGSPNSHVYQLLSPEDEGVEGRAFOVDPTSGSVTLGLVPLRAGQN 412
Qy 415 TPQYNLTIEVSKD--FKTLCFVQINVIDINDQIPIFEKSYGNLTLAEDTNIIGSTILT 472
Db 413 ILLVLVLAADLAGAGGFSSTCEVEVAVTDINDHAPEFITSQIGPISLPEDEPGTLVAML 472
```

```
Db 643 -----LTLAPVPSQ-----YLCTPRQDHGLIVGSPSKDPDLASGHGP-YSFT 683
Qy 707 LG-SGSLQNDWEVSKINGTHARLSTRHTDEERAYVVLIRINDGRPPLEIGVSLPVTFC 765
Db 684 LGPNPTVQRDWRQLTGLNSHAYLTALHWVEPREHIIIPVVVSHNAQ-----MWQLLVRIVC 740
Qy 766 SC-VEGSCFRPAGHOTGIPVGMAGILLTLLVIGILAVVFIKDKGKNVESAAQA 824
Db 741 RCNVEGQCMRKVRGMRKGMPTKLSAVGLVGLTVAIGIFLILIFTHWTMSRKKDPDQPADS 800
Qy 825 SEVK 828
Db 801 VPLK 804

RESULT 9
US-10-063-502-98
; Sequence 98, Application US/10063502
; Publication No. US20030023042A1
; GENERAL INFORMATION:
; APPLICANT: Eaton, Dan L.
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3230R1C1
; CURRENT APPLICATION NUMBER: US/10/063,502
; CURRENT FILING DATE: 2002-05-01
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 98
; LENGTH: 807
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-063-502-98

Query Match      21.0%; Score 908.5; DB 9; Length 807;
Best Local Similarity 30.2%; Pred. No. 1.6e-59;
Matches 237; Conservative 159; Mismatches 335; Indels 53; Gaps 21;

Qy 62 LTGE----TDNIFVIERE-GLLYYNRALDRETRSTHNLQVAALDANGIIIEGPPVPIIEV 116
Db 57 LSGDSGRKATEGPFAMDPDGSGLLVTRALDREEQAEYQLQVTFLEMQDGHVLMGPQVPLVHV 116
Qy 117 KDINDNRPFTLOSKEYSVRONSRPGPFLLVYNATDLDPATNGQLYQIVIQPLMINN 176
Db 117 KDENDQVPHFSQAIYRARLSRGTRPGIPFLFLEASDRDEPTANSDLRFLHLSQAQPS 176
Qy 177 VMYFOINNKTKAISLTREGSQELNPAKNSYNLVISKMGQSGSENSFSDTTSVDIIYTE 236
Db 177 PDMFOLEPRLGALALS PKGSTSLDHALERTYQLLVQVKMDGQA--SGHQATATVEVSII 235
Qy 237 NIWKAPRPVENVSTDPHPKIKITQVRWNPFGAQYSLVDKEKLPFRFFPSIDQEGDIYVTQ 296
Db 236 STWVSLEPIHLAENLKVLVPHHMAQVHWSGGDVHYHL---ESHPPGPFVNAEGLNYVTR 292
Qy 297 PLDREEDKDAYVYAVAKDEYCKPLSYPLEIHVKVDINDNPPTCPSPVTVEVOENERLG 356
Db 293 ELDRQAQAEYLLQVRAQNSHGDEYAAPLEHLVYMDENDNVPICPPRDPVTYSIPELSPG 352
Qy 357 NSIGTLTAHDEENTANSFLNYRIVEQTPKLPMDG-LFLIQTVAGMLQAKOSLKK-OD 414
Db 353 TEVTRLSAEDADAPGSPNSHVYQLLSPEDEGVEGRAFOVDPTSGSVTLGLVPLRAGQN 412
Qy 415 TPQYNLTIEVSKD--FKTLCFVQINVIDINDQIPIFEKSYGNLTLAEDTNIIGSTILT 472
Db 413 ILLVLVLAADLAGAGGFSSTCEVEVAVTDINDHAPEFITSQIGPISLPEDEPGTLVAML 472
```

```
QY 473 QATDAD-EPTGSSKIL-YHIKDSBGRGLGVDTDPHTNTGYVLIK--KPLDFETAAVSN 528
| | | | | : : : : | : | : | | : | : | : | : | : | : | : | : |
Db 473 TADADLEP---AFRLMDFAIRGDTGTEGLDWEF--DSGHVRLRLCKNLNLYEAAPSHE 527
| | | | | : : : : | : | : | | : | : | : | : | : | : | : | : |
QY 529 IVKAEINPEPLVGVKYKNASSFAKFTLITVDVNEAPQFSQHFQAKYSEDAIGTKYGNV 588
| : : | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 528 VVVVVQVAKLV-GPGGPGCATATVTLVERVMPPPKLDQESYEASVPISAPAGSFLLTI 586
| : : | | | : | : | : | : | : | : | : | : | : | : | : | : |
QY 589 TAKDPEGLDISYSLRGDTRGWLKIDHVTGEIFSVAPLD-REAGSPYRVQVVAVEVGSSSL 647
| | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 587 OPSDPISTRLRFSLVNDSEGMWCKIEKFSGEVHTAQSLQGAQGDGTITVLVEAQDTA---- 642
| : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 648 SSVSEPHILMDVNDNPRLLAKDVTGLFFCHPLSAPGSLIF-BATDDQHLFRGPHPTFS 706
| | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 643 -----LTLAPVPSQ-----YLCPTRODHGLIYSGPSKDPDLASGHCP-YSFT 683
| | | | | : | : | : | : | : | : | : | : | : | : | : | : |
QY 707 LG-SGSLQNDWEVSKINGTHARLSTRHTDDEERAYVYVLIINDDGRPPLEGIVSLPVTF 765
| : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 684 LGPNPTVQDWRQLQTLNGSHAYLTALHVVPEPREHIIPVVVSHNAQ-----MWQLLVRVIVC 740
| : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 766 SC-VEGSCFRPAGHGTGPIVGNVAGVLTLLTLLVIGILAVVFIRIKKDKGKNVESQAQ 824
| | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 741 RCNVEGQCMRKVRGKMGKMPKLSAVGLVGLVLAIGLIFLILFIETHWTMSRKKDPDPADS 800
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 825 SEVK 828
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 801 VPLK 804
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 10
US-10-227-884-188
; Sequence 188, Application US/10227884
; Publication No. US20030027988A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Desnoyers, Luc
; APPLICANT: Gerritsen, Mary
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Philippe F.
; APPLICANT: Watanabe, Colin L.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: F3530PIC79
; CURRENT APPLICATION NUMBER: US/10/227,884
; CURRENT FILING DATE: 2002-08-26
; PRIOR APPLICATION NUMBER: 10/119,480
; PRIOR FILING DATE: 2002-04-09
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/062287
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063549
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/064103
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/069873
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079294
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: 60/079656
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/079728
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/081819
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081955
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082804
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/084441
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/085323
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/086392
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089905
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090691
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/095302
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: 60/095318
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: 60/095916
; PRIOR FILING DATE: 1998-08-10
; PRIOR APPLICATION NUMBER: 60/096146
; PRIOR FILING DATE: 1998-08-11
; PRIOR APPLICATION NUMBER: 60/096791
; PRIOR FILING DATE: 1998-08-17
; PRIOR APPLICATION NUMBER: 60/097986
; PRIOR FILING DATE: 1998-08-26
; PRIOR APPLICATION NUMBER: 60/098544
; PRIOR FILING DATE: 1998-08-31
; PRIOR APPLICATION NUMBER: 60/099596
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099598
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099803
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099811
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099812
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099816
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/100038
; PRIOR FILING DATE: 1998-09-11
; PRIOR APPLICATION NUMBER: 60/100385
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: 60/100390
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: 60/100627
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100848
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/100919
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/101477
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101738
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101741
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101786
```





RESULT 11  
US-10-230-163-188  
; Sequence 189, Application US/10230163  
; Publication No. US20030036635A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Gerritsen, Mary  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, J. Christopher  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stephan, Jean-Philippe F.  
; APPLICANT: Watanabe, Colin L.  
; APPLICANT: Wood, William I.  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: F3530P1C96  
; CURRENT APPLICATION NUMBER: US/10/230,163  
; CURRENT FILING DATE: 2002-08-28  
; PRIOR APPLICATION NUMBER: 10/119,480  
; PRIOR FILING DATE: 2002-04-09  
; PRIOR APPLICATION NUMBER: 60/059113  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/062287  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/063549  
; PRIOR FILING DATE: 1997-10-28  
; PRIOR APPLICATION NUMBER: 60/064103  
; PRIOR FILING DATE: 1997-10-31  
; PRIOR APPLICATION NUMBER: 60/069873  
; PRIOR FILING DATE: 1997-12-17  
; PRIOR APPLICATION NUMBER: 60/078910  
; PRIOR FILING DATE: 1998-03-20  
; PRIOR APPLICATION NUMBER: 60/079294  
; PRIOR FILING DATE: 1998-03-25  
; PRIOR APPLICATION NUMBER: 60/079656  
; PRIOR FILING DATE: 1998-03-26  
; PRIOR APPLICATION NUMBER: 60/079728  
; PRIOR FILING DATE: 1998-03-27  
; PRIOR APPLICATION NUMBER: 60/081819  
; PRIOR FILING DATE: 1998-04-15  
; PRIOR APPLICATION NUMBER: 60/081955  
; PRIOR FILING DATE: 1998-04-15  
; PRIOR APPLICATION NUMBER: 60/082804  
; PRIOR FILING DATE: 1998-04-22  
; PRIOR APPLICATION NUMBER: 60/084441  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: 60/085323  
; PRIOR FILING DATE: 1998-05-13  
; PRIOR APPLICATION NUMBER: 60/085579  
; PRIOR FILING DATE: 1998-05-15  
; PRIOR APPLICATION NUMBER: 60/086392  
; PRIOR FILING DATE: 1998-05-22  
; PRIOR APPLICATION NUMBER: 60/089532  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089538  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089905  
; PRIOR FILING DATE: 1998-06-18  
; PRIOR APPLICATION NUMBER: 60/090472  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090557  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090691  
; PRIOR FILING DATE: 1998-06-25  
; PRIOR APPLICATION NUMBER: 60/090695  
; PRIOR FILING DATE: 1998-06-25  
; PRIOR APPLICATION NUMBER: 60/091982  
; PRIOR FILING DATE: 1998-07-07  
; PRIOR APPLICATION NUMBER: 60/095302  
; PRIOR FILING DATE: 1998-08-04  
; PRIOR APPLICATION NUMBER: 60/095318  
; PRIOR FILING DATE: 1998-08-04  
; PRIOR APPLICATION NUMBER: 60/095916  
; PRIOR FILING DATE: 1998-08-10  
; PRIOR APPLICATION NUMBER: 60/096146  
; PRIOR FILING DATE: 1998-08-11  
; PRIOR APPLICATION NUMBER: 60/096791  
; PRIOR FILING DATE: 1998-08-17  
; PRIOR APPLICATION NUMBER: 60/097986  
; PRIOR FILING DATE: 1998-08-26  
; PRIOR APPLICATION NUMBER: 60/098544  
; PRIOR FILING DATE: 1998-08-31  
; PRIOR APPLICATION NUMBER: 60/099596  
; PRIOR FILING DATE: 1998-09-09  
; PRIOR APPLICATION NUMBER: 60/099598  
; PRIOR FILING DATE: 1998-09-09  
; PRIOR APPLICATION NUMBER: 60/099803  
; PRIOR FILING DATE: 1998-09-10  
; PRIOR APPLICATION NUMBER: 60/099811  
; PRIOR FILING DATE: 1998-09-10  
; PRIOR APPLICATION NUMBER: 60/099812  
; PRIOR FILING DATE: 1998-09-10  
; PRIOR APPLICATION NUMBER: 60/099816  
; PRIOR FILING DATE: 1998-09-10  
; PRIOR APPLICATION NUMBER: 60/100038  
; PRIOR FILING DATE: 1998-09-11  
; PRIOR APPLICATION NUMBER: 60/100385  
; PRIOR FILING DATE: 1998-09-15  
; PRIOR APPLICATION NUMBER: 60/100390  
; PRIOR FILING DATE: 1998-09-15  
; PRIOR APPLICATION NUMBER: 60/100627  
; PRIOR FILING DATE: 1998-09-16  
; PRIOR APPLICATION NUMBER: 60/100848  
; PRIOR FILING DATE: 1998-09-18  
; PRIOR APPLICATION NUMBER: 60/100919  
; PRIOR FILING DATE: 1998-09-17  
; PRIOR APPLICATION NUMBER: 60/101477  
; PRIOR FILING DATE: 1998-09-23  
; PRIOR APPLICATION NUMBER: 60/101738  
; PRIOR FILING DATE: 1998-09-24  
; PRIOR APPLICATION NUMBER: 60/101741  
; PRIOR FILING DATE: 1998-09-24  
; PRIOR APPLICATION NUMBER: 60/101786  
; PRIOR FILING DATE: 1998-09-25  
; PRIOR APPLICATION NUMBER: 60/101916  
; PRIOR FILING DATE: 1998-09-24  
; PRIOR APPLICATION NUMBER: 60/101922  
; PRIOR FILING DATE: 1998-09-24  
; PRIOR APPLICATION NUMBER: 60/106178  
; PRIOR FILING DATE: 1998-10-28  
; PRIOR APPLICATION NUMBER: 60/106248  
; PRIOR FILING DATE: 1998-10-29  
; PRIOR APPLICATION NUMBER: 60/106464  
; PRIOR FILING DATE: 1998-10-30  
; PRIOR APPLICATION NUMBER: 60/106905  
; PRIOR FILING DATE: 1998-11-03  
; PRIOR APPLICATION NUMBER: 60/108787  
; PRIOR FILING DATE: 1998-11-17  
; PRIOR APPLICATION NUMBER: 60/108801  
; PRIOR FILING DATE: 1998-11-17  
; PRIOR APPLICATION NUMBER: 60/108849  
; PRIOR FILING DATE: 1998-11-18  
; PRIOR APPLICATION NUMBER: 60/112422  
; PRIOR FILING DATE: 1998-12-15  
; PRIOR APPLICATION NUMBER: 60/113296  
; PRIOR FILING DATE: 1998-12-22  
; PRIOR APPLICATION NUMBER: 60/113605  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/113621  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/115558





Db 413 ILLVLAMDLAGAEFGFSTCEVEAVTNDINDHAPEFITSQIGPISLPEDVEPGTLVAML 472  
Qy 473 QATDAD-EPFTGSSKIL-YHIKGDSEGLGVDPDHTNTGYVIK--KPLDFFETAASN 528  
Db 473 TAIDADLEP---AFRLMDFAIERGDTGTFGLDWEF--DSGHVRLRLCKNLSYEAAPSHE 527  
Qy 529 IVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFOAKYSEDVAIGTKVGNV 588  
Db 528 VVVVQSVAKLV-GPGPGGATATVTLVERVMPKLDQESYEASVPISAPAGSFLITI 586  
Qy 589 TAKDPEGLDTSYLRGDTGRWLKIDHVTGEIFSVAPLD-REAGSPYRVQVVAEVEGSSSL 647  
Db 587 QPSDPISRTRLRFSLVNDSEGWLCEKFSGEVHTAQSLQGAQPGDTYTLVLEAQDTA---- 642  
Qy 648 SSVSEFHLILMDVNDNPRLAKDYTGFFCHPLSAPGSLIF-EATDDDDQHLFRGPHFTFS 706  
Db 643 -----LTLAPVPSQ-----YLCPTRODHGLIVSGPSKDPDLASGHP-YSFT 683  
Qy 707 LG-SGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGRPPLEGIVSLPVTFC 765  
Db 684 LGPNPTVQDORWRLQTLNGSHAYLTLALHWEVPREHIIIPVVVSHNAQ---MWQLLVRVIVC 740  
Qy 766 SC-VEGSCFRPAGHOTGIPVTGMAVGILLTLLVIGIILAVVFIRIKDKGKNVESQA 824  
Db 741 RCNVEGQCMRKVRGMKGMPTKLSAVGLVGLTVAIGIFLILIFHTWMSRKKDPQPADS 800  
Qy 825 SEVK 828  
Db 801 VPLK 804

## RESULT 14

US-10-230-338-188  
; Sequence 188, Application US/10230338  
; Publication No. US2003004934A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Gerritsen, Mary  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, J. Christopher  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stephan, Jean-Philippe F.  
; APPLICANT: Watanabe, Colin L.  
; APPLICANT: Wood, William I.  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
; FILE REFERENCE: P3530PIC92  
; CURRENT APPLICATION NUMBER: US/10/230,338  
; CURRENT FILING DATE: 2002-08-28  
; PRIOR APPLICATION NUMBER: 10/119,480  
; PRIOR FILING DATE: 2002-04-09  
; PRIOR APPLICATION NUMBER: 60/059113  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/062287  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/063549  
; PRIOR FILING DATE: 1997-10-28  
; PRIOR APPLICATION NUMBER: 60/064103  
; PRIOR FILING DATE: 1997-10-31  
; PRIOR APPLICATION NUMBER: 60/069873  
; PRIOR FILING DATE: 1997-12-17  
; PRIOR APPLICATION NUMBER: 60/078910  
; PRIOR FILING DATE: 1998-03-20  
; PRIOR APPLICATION NUMBER: 60/079294  
; PRIOR FILING DATE: 1998-03-25  
; PRIOR APPLICATION NUMBER: 60/079656  
; PRIOR FILING DATE: 1998-03-26  
; PRIOR APPLICATION NUMBER: 60/079728  
; PRIOR FILING DATE: 1998-03-27  
; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 246  
; SEQ ID NO 188  
; LENGTH: 807  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-230-338-188

Query Match 21.0%; Score 908.5; DB 9; Length 807;  
Best Local Similarity 30.2%; Pred. No. 1.6e-59;  
Matches 237; Conservative 159; Mismatches 335; Indels 53; Gaps 21;

Qy 62 LTGE-----TNIFVIERE-GLLYNNRDLRETRTHNLQVAALDANGILVGGPPIIEV 116  
Db 57 LSGDSGKATEGPFAMDPDSGFLLVTRALDREEAQYQLVTEMQDGHVLMGPQVILHV 116  
Qy 117 KDINDNPTFLQSKYEGSRQNSRPKFLYVYNATDDPATNGQLYXOYIOLPMINN 176  
Db 117 KDENDQPHFESQAIYRLARLSRGTRPGIPFLFLEASDRDEPTANSDLRFHLSQAPQS 176  
Qy 177 VMYQINNKTKGALSUTREGSOELNPAKNPSYLVISVKMDGGQSGNSFSDTTSYDIIVTE 236  
Db 177 PMFQLEPRLGALALSPKGSTSLHALERTYQLLVQVKMGDQA-SGHQATATVEVSIE 235  
Qy 237 NIWKAPKPVEMVENSTDPHPKITQVRWNPQAOYSLVDKEKLPFRFESIDQGDIIYVQ 296  
Db 236 STWYSLEPIHLAENKLVLYPHHMAQVHWSGGDVHYHL----ESHPPGPFVEVNAENLYYTR 292  
Qy 297 PLDREKDAYVYAVAKDEYKGLPSYPLEIHHVKVDINDNPPTCPSPVTVFEVENERLG 356  
Db 293 ELDRQAQAEYLLQVRAQNSHGEDIAAPLEHLVYDENDNVPICPRPDPTVISIELSPPG 352  
Qy 357 NSIGTLTAHDRDEENTANSFLNYRIVEQTQKLPMDG-LFLIQTYYAGMLQAKQSLKK-QD 414  
Db 353 TEVTRLAEDADAPGSPNSHVYQLLSPEDGVEGRAFOVDPTSGSVTLGLVPLRAQON 412  
Qy 415 TPQYNLTIEVSKD--FKTLCFVQINVIDINDQIPIEKSQDYNLTAEADNIGSTILTI 472  
Db 413 ILLVLAMDLAGAEFGFSTCEVEAVTNDINDHAPEFITSQIGPISLPEDVEPGTLVAML 472  
Qy 473 QATDAD-EPFTGSSKIL-YHIKGDSEGLGVDPDHTNTGYVIK--KPLDFFETAASN 528  
Db 473 TAIDADLEP---AFRLMDFAIERGDTGTFGLDWEF--DSGHVRLRLCKNLSYEAAPSHE 527  
Qy 529 IVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFOAKYSEDVAIGTKVGNV 588  
Db 528 VVVVQSVAKLV-GPGPGGATATVTLVERVMPKLDQESYEASVPISAPAGSFLITI 586  
Qy 589 TAKDPEGLDTSYLRGDTGRWLKIDHVTGEIFSVAPLD-REAGSPYRVQVVAEVEGSSSL 647  
Db 587 QPSDPISRTRLRFSLVNDSEGWLCEKFSGEVHTAQSLQGAQPGDTYTLVLEAQDTA---- 642  
Qy 648 SSVSEFHLILMDVNDNPRLAKDYTGFFCHPLSAPGSLIF-EATDDDDQHLFRGPHFTFS 706  
Db 643 -----LTLAPVPSQ-----YLCPTRODHGLIVSGPSKDPDLASGHP-YSFT 683  
Qy 707 LG-SGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGRPPLEGIVSLPVTFC 765  
Db 684 LGPNPTVQDORWRLQTLNGSHAYLTLALHWEVPREHIIIPVVVSHNAQ---MWQLLVRVIVC 740  
Qy 766 SC-VEGSCFRPAGHOTGIPVTGMAVGILLTLLVIGIILAVVFIRIKDKGKNVESQA 824  
Db 741 RCNVEGQCMRKVRGMKGMPTKLSAVGLVGLTVAIGIFLILIFHTWMSRKKDPQPADS 800  
Qy 825 SEVK 828  
Db 801 VPLK 804

## RESULT 15

US-10-063-518-98  
; Sequence 98, Application US/10063518  
; Publication No. US20030049735A1  
; GENERAL INFORMATION:

```

; APPLICANT: Eaton, Dan L.
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3230R1C1
; CURRENT APPLICATION NUMBER: US/10/063,518
; CURRENT FILING DATE: 2002-05-01
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 98
; LENGTH: 807
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-10-063-518-98

Query Match      21.0%; Score 908.5; DB 9; Length 807;
Best Local Similarity 30.2%; Pred. No. 1.6e-59;
Matches 237; Conservative 159; Mismatches 335; Indels 53; Gaps 21;

Qy 62 LTGE---TDNIFVIERE-GLLYNNRDLRETRSTHNLQVAALDANGIIIVEGVPVITIEV 116
Db 57 LSGDSKATGEPFAMPDPSGFLLVTRALDREEQAEYQLQVLEMDQGHVWGQPVLVHV 116
Qy 117 KDINDNRPTFLQSKYEGSVQRNRPKPFLLYVYNATLDDPATNGQLYQYIVIQLPMINN 176
Db 117 KDENDQVPHEFSQAIYRRLSRGTRPGIPFLLEASDRDEPGTANSDLRPHILSQAPQPS 176
Qy 177 VMYFOINNKGTGATSLREGSOELNPAKNPSYLVISVKMDGQSENSEFSDTTSVDIIVTE 236
Db 177 PMFQLEPRLGALALSPKSTSDHALERTYQLLVQVKMDGQA-SGHQATATVEVSIIIE 235
Qy 237 NIWKAPKPVEMVNSTDPHPKIKITQVRNDPQGAQYSLVDKEKLPFRPFISIDQEGDIYVTQ 296
Db 236 STWVSLEPIHIAENLKVLYPHHMAQVHWSGGDVHYHL---ESHPPGPFVEVNAEGNLVTR 292
Qy 297 PLDREKDAYVFAKDEYKGLSPLEIYHVKVKDINDNPPTCPSPVTVEVOENRERG 356
Db 293 ELDRQAQAEYLLQVRAQNSHGDEYAAPLEHLVLMENDENVVICPPRDPPTVSIPELSPPG 352
Qy 357 NSIGTLTAHDDRDEANTANFLNVRIVEQTPKLPMDG-LFLIQTIVAGMLQAKOSLKK-QD 414
Db 353 TEVTRLSAEDADAPGSPNSHVYQLLSPEPEDGVEGRAFOVDPSTGSGVTGLVPLPRAGQN 412
Qy 415 TPOYNLTIEVSKD--FKTLCFVQINVIDINDQIPIFEKSDYGNLTALAEEDTNIIGSTILTI 472
Db 413 ILLVLAMDLAGAEGGFSSTCEVEAVTDINDHAPEFITSQIGPISLPEDVEPGTLVAML 472
Qy 473 QATDAD-EPTTGSKIL-YHIIKGDSEGRGLGVDTDPHTNTGYVIK--KPLDFTAAVSN 528
Db 473 TADADLEP---AFRLMDFAIERDTEGTGLDWEV--DSGHVRLRCKNLSEYAAPSHSHE 527
Qy 529 IVFKAENPEPLVGVKYNASSFAKFTLIVTDVNEAPQFSQHVQAKYSEDVAIGTKVGNV 588
Db 528 VVVVQSVAKLV-GPGGPGATATVTLVERVMPKPKLDQESYEASVPIAPAGSFLTTI 586
Qy 589 TAKDPEGLDISYLRGDTGRGLKIDHVTGIFSVAPLD-REAGSPYRVQVQVATEVGVSSSL 647
Db 587 QPSDPISTLRFSLVNDSEGNLCIEKESGEVHTAQSLQGAQPGDTYTVLVEAQDTA---- 642
Qy 648 SSVSEPHLIIMLVNDNPRLAKDYTGTFCHPLSAPGSLIP-EATDDQHLFRGPHFTFS 706
Db 643 -----LTLAPVPSQ-----YLCPTQDHLIVSGPSKPDLASGHGP-YSET 683
Qy 707 LG-SGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFC 765
Db 684 LGPNPTVQDNRQLQTLNGSHAYLTALHWVEPREHIIPVVVVSHNAQ----MWQLLVRVIVC 740

```

```

Qy 766 SC-VEGSCFRPAGHQGTGIPVGMVAGILLTLLVIGIILAVVFIKIKDKGKDNVESQAQ 824
Db 741 RCNVEGQCMRKVGRMGKMPKLSAVGILVGTVAIGIFLILFIHTMTMSRKKDQDPADS 800
Qy 825 SEVK 828
Db 801 VPLK 804

Search completed: April 2, 2003, 16:05:40
Job time : 46.7397 secs

```

**THIS PAGE BLANK (USPIC,**

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 2, 2003, 15:52:30 ; Search time 436.895 Seconds  
(without alignments)  
1227.796 Million cell updates/sec

Title: US-09-079-678-178  
Perfect score: 4321  
Sequence: 1 MILOAHLHSLCLMLYLATG.....DKGKDNVESAQAEVKPLRS 832

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues  
Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main:\*

1: /cgn2\_6/ptodata/1/paa/PCTUS\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/paa/US06\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/paa/US07\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/paa/US08\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/paa/US081\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/paa/US082\_COMB.pep.\*  
7: /cgn2\_6/ptodata/1/paa/US083\_COMB.pep.\*  
8: /cgn2\_6/ptodata/1/paa/US084\_COMB.pep.\*  
9: /cgn2\_6/ptodata/1/paa/US085\_COMB.pep.\*  
10: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep.\*  
11: /cgn2\_6/ptodata/1/paa/US087\_COMB.pep.\*  
12: /cgn2\_6/ptodata/1/paa/US088\_COMB.pep.\*  
13: /cgn2\_6/ptodata/1/paa/US089\_COMB.pep.\*  
14: /cgn2\_6/ptodata/1/paa/US090\_COMB.pep.\*  
15: /cgn2\_6/ptodata/1/paa/US091\_COMB.pep.\*  
16: /cgn2\_6/ptodata/1/paa/US092\_COMB.pep.\*  
17: /cgn2\_6/ptodata/1/paa/US093\_COMB.pep.\*  
18: /cgn2\_6/ptodata/1/paa/US094\_COMB.pep.\*  
19: /cgn2\_6/ptodata/1/paa/US095\_COMB.pep.\*  
20: /cgn2\_6/ptodata/1/paa/US096\_COMB.pep.\*  
21: /cgn2\_6/ptodata/1/paa/US097\_COMB.pep.\*  
22: /cgn2\_6/ptodata/1/paa/US098\_COMB.pep.\*  
23: /cgn2\_6/ptodata/1/paa/US099\_COMB.pep.\*  
24: /cgn2\_6/ptodata/1/paa/US100\_COMB.pep.\*  
25: /cgn2\_6/ptodata/1/paa/US101\_COMB.pep.\*  
26: /cgn2\_6/ptodata/1/paa/US102\_COMB.pep.\*  
27: /cgn2\_6/ptodata/1/paa/US60\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES									
Result No.	Query				Description				
	Score	Match	Length	DB	ID				
1	4321	100.0	832	1	PCT-US98-10088-178	Sequence 178, App			
2	4321	100.0	832	4	US-08-013-462B-1	Sequence 1, Appli			
3	4321	100.0	832	14	US-09-079-678-178	Sequence 178, App			
4	4321	100.0	832	14	US-09-079-678-178	Sequence 178, App			
5	4321	100.0	832	14	US-09-079-723-178	Sequence 178, App			
6	4321	100.0	832	14	US-09-079-723-178	Sequence 178, App			

7	4321	100.0	832	14	US-09-079-819-178		Sequence 178, App
8	4321	100.0	832	14	US-09-079-819-178		Sequence 178, App
9	4321	100.0	832	20	US-09-609-448A-1081		Sequence 1081, Ap
10	4321	100.0	832	20	US-09-649-811-1081		Sequence 1081, Ap
11	4321	100.0	832	21	US-09-791-537-48350		Sequence 48350, A
12	4321	100.0	832	22	US-09-833-263-1081		Sequence 1081, Ap
13	4321	100.0	832	23	US-09-922-217-1081		Sequence 1081, Ap
14	4321	100.0	832	23	US-09-443-780C-15		Sequence 15, Appl
15	4321	100.0	832	24	US-10-025-380-1081		Sequence 1081, Ap
16	4321	100.0	832	25	US-10-174-677-16		Sequence 16, Appl
17	4317	99.9	832	1	PCT-US02-27737-131		Sequence 131, Appl
18	4303	99.6	832	21	US-09-791-537-75515		Sequence 75515, A
19	4208	97.4	958	22	US-09-833-263-1087		Sequence 1087, Ap
20	4208	97.4	958	23	US-09-922-217-1087		Sequence 1087, Ap
21	4208	97.4	958	24	US-10-025-380-1087		Sequence 1087, Ap
22	3450	79.8	827	21	US-09-791-537-1000		Sequence 1000, Ap
23	3414	79.0	827	21	US-09-791-537-53390		Sequence 53390, A
24	2332	54.0	578	21	US-09-791-537-149885		Sequence 149885,
25	2328	53.9	578	21	US-09-791-537-152759		Sequence 152759,
26	954	22.1	829	21	US-09-791-537-44940		Sequence 44940, A
27	922.5	21.3	829	21	US-09-791-537-75511		Sequence 75511, A
28	922.5	21.3	829	25	US-10-174-677-15		Sequence 15, Appl
29	908.5	21.0	807	23	US-09-946-374-229		Sequence 229, App
30	908.5	21.0	807	24	US-10-006-041A-229		Sequence 229, App
31	908.5	21.0	807	24	US-10-006-063A-229		Sequence 229, App
32	908.5	21.0	807	24	US-10-006-116A-229		Sequence 229, App
33	908.5	21.0	807	24	US-10-006-117A-229		Sequence 229, App
34	908.5	21.0	807	24	US-10-006-130A-229		Sequence 229, App
35	908.5	21.0	807	24	US-10-006-172A-229		Sequence 229, App
36	908.5	21.0	807	24	US-10-006-485A-229		Sequence 229, App
37	908.5	21.0	807	24	US-10-006-746A-229		Sequence 229, App
38	908.5	21.0	807	24	US-10-006-768A-229		Sequence 229, App
39	908.5	21.0	807	24	US-10-006-818A-229		Sequence 229, App
40	908.5	21.0	807	24	US-10-006-856A-229		Sequence 229, App
41	908.5	21.0	807	24	US-10-006-867-98		Sequence 98, Appl
42	908.5	21.0	807	24	US-10-007-194A-229		Sequence 229, App
43	908.5	21.0	807	24	US-10-007-236A-229		Sequence 229, App
44	908.5	21.0	807	24	US-10-011-671A-229		Sequence 229, App
45	908.5	21.0	807	24	US-10-011-692A-229		Sequence 229, App

ALIGNMENTS

RESULT 1  
PCT-US98-10088-178  
; Sequence 178, Application PC/TUS9810088  
; GENERAL INFORMATION:  
; APPLICANT: CYTOGEN CORPORATION  
; APPLICANT: ELAN CORPORATION, PLC  
; TITLE OF INVENTION: RANDOM PEPTIDES THAT BIND TO GASTRO-  
; TITLE OF INVENTION: INTESTINAL TRACT (GIT) TRANSPORT RECEPTORS AND RELATED METH  
; NUMBER OF SEQUENCES: 265  
; CORRESPONDENCE ADDRESS: 265  
; ADDRESSEE: Pennie & Edmonds LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ Version 2.0  
; CURRENT APPLICATION DATA: PCT/US98/10088  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Misrock, S. Leslie  
; REGISTRATION NUMBER: 18,972  
; REFERENCE/DOCKET NUMBER: 1101-209

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 178:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 832 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
PCT-US98-10088-178

Query Match 100.0%; Score 4321; DB 1; Length 832;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 832; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MILOAHLHSCLLMYLYLATYGGQEGKFSGLPKMTFSIYEGQEPSQIIFQKAPPAVTF 60  
DB 1 MILOAHLHSCLLMYLYLATYGGQEGKFSGLPKMTFSIYEGQEPSQIIFQKAPPAVTF 60

QY 61 ELTGEDTNIIEVEGGLLYNRLDRETRSTHNLQVAALDANGIIIVGPPITIEVKDIN 120  
DB 61 ELTGEDTNIIEVEGGLLYNRLDRETRSTHNLQVAALDANGIIIVGPPITIEVKDIN 120

QY 121 DNRPTFLQSKYEGSVRQNSRPGKFLYVNATDLDPATNGQLYQIYIQLPMINNNMYF 180  
DB 121 DNRPTFLQSKYEGSVRQNSRPGKFLYVNATDLDPATNGQLYQIYIQLPMINNNMYF 180

QY 181 QINNKTAISLTREGSQELNPAKNPSYLVSKDMGQSGENSEFSDTTSVDIIIVTENIWK 240  
DB 181 QINNKTAISLTREGSQELNPAKNPSYLVSKDMGQSGENSEFSDTTSVDIIIVTENIWK 240

QY 241 AKPVEVVENSTDPHPKIKITQVRWNPDAQYSLVDKELPRFPESIDQEGDIYVTPQLDR 300  
DB 241 AKPVEVVENSTDPHPKIKITQVRWNPDAQYSLVDKELPRFPESIDQEGDIYVTPQLDR 300

QY 301 EEKDAYVYFAVAKDEYKPLSYPLEIHVKVDINDNPPTCSPVTVEVQENERLGNISG 360  
DB 301 EEKDAYVYFAVAKDEYKPLSYPLEIHVKVDINDNPPTCSPVTVEVQENERLGNISG 360

QY 361 TLTAHDEENTANSFLNYRIVEQTPKLPMDGLFLIQTAYAGMLQAKQSLKQDTPQYNL 420  
DB 361 TLTAHDEENTANSFLNYRIVEQTPKLPMDGLFLIQTAYAGMLQAKQSLKQDTPQYNL 420

QY 421 TIEVSDKDKFTLCFVQINVIDINDOIPFEKSDYGNLTAEADTNGISTILTIOATDADEP 480  
DB 421 TIEVSDKDKFTLCFVQINVIDINDOIPFEKSDYGNLTAEADTNGISTILTIOATDADEP 480

QY 481 FTGSSKILYHIKGDSEGRGLGVDTDPHTNTGYVLIKKPLDFETAASVNIIVFKAENPEPLV 540  
DB 481 FTGSSKILYHIKGDSEGRGLGVDTDPHTNTGYVLIKKPLDFETAASVNIIVFKAENPEPLV 540

RESULT 2

US-08-013-462B-1  
Sequence 1, Application US/08013462B  
GENERAL INFORMATION:  
APPLICANT: Anne H. Dantzig, et al.  
TITLE OF INVENTION: Mammalian Influx Peptide  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Eli Lilly and Company  
STREET: Lilly Corporate Center  
CITY: Indianapolis  
STATE: IN  
COUNTRY: U.S.A.  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.0 Mb storage  
COMPUTER: Macintosh  
OPERATING SYSTEM: Macintosh  
SOFTWARE: Microsoft Word  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/013,462B  
FILING DATE: 19930204  
CLASSIFICATION: 435  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 832 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-013-462B-1

Query Match 100.0%; Score 4321; DB 4; Length 832;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 832; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MILOAHLHSCLLMYLYLATYGGQEGKFSGLPKMTFSIYEGQEPSQIIFQKAPPAVTF 60  
DB 1 MILOAHLHSCLLMYLYLATYGGQEGKFSGLPKMTFSIYEGQEPSQIIFQKAPPAVTF 60

QY 61 ELTGEDTNIIEVEGGLLYNRLDRETRSTHNLQVAALDANGIIIVGPPITIEVKDIN 120  
DB 61 ELTGEDTNIIEVEGGLLYNRLDRETRSTHNLQVAALDANGIIIVGPPITIEVKDIN 120

QY 121 DNRPTFLQSKYEGSVRQNSRPGKFLYVNATDLDPATNGQLYQIYIQLPMINNNMYF 180  
DB 121 DNRPTFLQSKYEGSVRQNSRPGKFLYVNATDLDPATNGQLYQIYIQLPMINNNMYF 180

QY 181 QINNKTAISLTREGSQELNPAKNPSYLVSKDMGQSGENSEFSDTTSVDIIIVTENIWK 240  
DB 181 QINNKTAISLTREGSQELNPAKNPSYLVSKDMGQSGENSEFSDTTSVDIIIVTENIWK 240

QY 241 AKPVEVVENSTDPHPKIKITQVRWNPDAQYSLVDKELPRFPESIDQEGDIYVTPQLDR 300  
DB 241 AKPVEVVENSTDPHPKIKITQVRWNPDAQYSLVDKELPRFPESIDQEGDIYVTPQLDR 300

QY 301 EEKDAYVYFAVAKDEYKPLSYPLEIHVKVDINDNPPTCSPVTVEVQENERLGNISG 360  
DB 301 EEKDAYVYFAVAKDEYKPLSYPLEIHVKVDINDNPPTCSPVTVEVQENERLGNISG 360

QY 361 TLTAHDEENTANSFLNYRIVEQTPKLPMDGLFLIQTAYAGMLQAKQSLKQDTPQYNL 420  
DB 361 TLTAHDEENTANSFLNYRIVEQTPKLPMDGLFLIQTAYAGMLQAKQSLKQDTPQYNL 420

QY 421 TIEVSDKDKFTLCFVQINVIDINDOIPFEKSDYGNLTAEADTNGISTILTIOATDADEP 480  
DB 421 TIEVSDKDKFTLCFVQINVIDINDOIPFEKSDYGNLTAEADTNGISTILTIOATDADEP 480

QY 481 FTGSSKILYHIKGDSEGRGLGVDTDPHTNTGYVLIKKPLDFETAASVNIIVFKAENPEPLV 540  
DB 481 FTGSSKILYHIKGDSEGRGLGVDTDPHTNTGYVLIKKPLDFETAASVNIIVFKAENPEPLV 540



QY 541 FGKYNASSFAKFTLIIVTDVNEAPQFSQHVFAQKSEDAVIGTKVGNVTAQKDPGLDISY 600  
Db 541 FGKYNASSFAKFTLIIVTDVNEAPQFSQHVFAQKSEDAVIGTKVGNVTAQKDPGLDISY 600  
QY 601 SLRGDTGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVADEVGGSSLSVSEFHLILMDV 660  
Db 601 SLRGDTGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVADEVGGSSLSVSEFHLILMDV 660  
QY 661 NDNPRLAKDYTLGFFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEYSK 720  
Db 661 NDNPRLAKDYTLGFFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEYSK 720  
QY 721 INGTARLSTRHTDFEERAYVVLIRINDGRRPPLLEGIVSLPVTFCSCVEGSCFRPAGHQT 780  
Db 721 INGTARLSTRHTDFEERAYVVLIRINDGRRPPLLEGIVSLPVTFCSCVEGSCFRPAGHQT 780  
QY 781 GIPVGMAGVILLTLLVIGIILAVVFIKDKGKDNVESQAQASEVKPLRS 832  
Db 781 GIPVGMAGVILLTLLVIGIILAVVFIKDKGKDNVESQAQASEVKPLRS 832

RESULT 3

US-09-079-678-178 Application US/09079678

GENERAL INFORMATION:

APPLICANT: Alvarez, Vernon L.  
APPLICANT: O'Mahony, Daniel J.  
APPLICANT: Lambkin, Imelda J.  
APPLICANT: Singleton, Judith  
APPLICANT: Patterson, Catherine A.  
APPLICANT: Cagney, Gerard M.  
APPLICANT: Belinka, Benjamin A.

ATTORNEY/AGENT INFORMATION:  
NAME: Mirock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 1101-220

TELEPHONE: 212-790-9090

TELEFAX: 212-869-9741

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 178:

SEQUENCE CHARACTERISTICS:

LENGTH: 832 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: unknown

MOLECULE TYPE: protein

US-09-079-678-178

Query Match 100.0%; Score 4321; DB 14; Length 832;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 832; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MILQAHLSLCLMLYLATGYQEGKFSGLPKMTFSIYEGQEPSQIIFQKANPPAVTF 60  
Db 1 MILQAHLSLCLMLYLATGYQEGKFSGLPKMTFSIYEGQEPSQIIFQKANPPAVTF 60  
QY 61 ELTGETONIFVIERGLLYNRAIDRETRSHNLQVAALDANGIIEGVPVITIEVKDIN 120  
Db 61 ELTGETONIFVIERGLLYNRAIDRETRSHNLQVAALDANGIIEGVPVITIEVKDIN 120  
QY 121 DNRRTFLQSKYEGSVRONSRPGKPLFYVNATDLDPPATNGQLYYQIIVQLPMNNMYF 180  
Db 121 DNRRTFLQSKYEGSVRONSRPGKPLFYVNATDLDPPATNGQLYYQIIVQLPMNNMYF 180  
QY 181 QINNKTAISLTRGSELNPAKNPSYNLVTSVKDMGQSGNSFSDTTSVDIIVTENIWK 240  
Db 181 QINNKTAISLTRGSELNPAKNPSYNLVTSVKDMGQSGNSFSDTTSVDIIVTENIWK 240  
QY 241 APKPVEMVNSTDPHPKITQVRWNPDCAOYSLVDKEKLPFPFSDIOEGDIYVTPQLDR 300  
Db 241 APKPVEMVNSTDPHPKITQVRWNPDCAOYSLVDKEKLPFPFSDIOEGDIYVTPQLDR 300  
QY 301 EKDAYSFVAVAKDEYKPLSYPLEIHVKVDINDNPPTCPSPVTVFEVQENRGLNSIG 360  
Db 301 EKDAYSFVAVAKDEYKPLSYPLEIHVKVDINDNPPTCPSPVTVFEVQENRGLNSIG 360  
QY 361 TLTAHDEENTANSFLNYRIVEQTPKLPMDGLFLIQTAYAGMLQAKSLKKQDTPQYNL 420  
Db 361 TLTAHDEENTANSFLNYRIVEQTPKLPMDGLFLIQTAYAGMLQAKSLKKQDTPQYNL 420  
QY 421 TIEVSDKDFKLCFQVINVIDINDQIPIFEKSDYGNLTAEEDTNGISTILITQATDADEP 480  
Db 421 TIEVSDKDFKLCFQVINVIDINDQIPIFEKSDYGNLTAEEDTNGISTILITQATDADEP 480  
QY 481 FTGSSKILYHIKGDSEGRGLGVDTPHTNTGYIHKPLDEETAASVNIIVFKAENPEPLV 540  
Db 481 FTGSSKILYHIKGDSEGRGLGVDTPHTNTGYIHKPLDEETAASVNIIVFKAENPEPLV 540  
QY 541 FGKYNASSFAKFTLIIVTDVNEAPQFSQHVFAQKSEDAVIGTKVGNVTAQKDPGLDISY 600  
Db 541 FGKYNASSFAKFTLIIVTDVNEAPQFSQHVFAQKSEDAVIGTKVGNVTAQKDPGLDISY 600  
QY 601 SLRGDTGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVADEVGGSSLSVSEFHLILMDV 660  
Db 601 SLRGDTGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVADEVGGSSLSVSEFHLILMDV 660  
QY 661 NDNPRLAKDYTLGFFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEYSK 720  
Db 661 NDNPRLAKDYTLGFFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEYSK 720  
QY 721 INGTARLSTRHTDFEERAYVVLIRINDGRRPPLLEGIVSLPVTFCSCVEGSCFRPAGHQT 780  
Db 721 INGTARLSTRHTDFEERAYVVLIRINDGRRPPLLEGIVSLPVTFCSCVEGSCFRPAGHQT 780  
QY 781 GIPVGMAGVILLTLLVIGIILAVVFIKDKGKDNVESQAQASEVKPLRS 832  
Db 781 GIPVGMAGVILLTLLVIGIILAVVFIKDKGKDNVESQAQASEVKPLRS 832

RESULT 4

US-09-079-678-178

Sequence 178, Application US/09079678A

GENERAL INFORMATION:

APPLICANT: Alvarez, Vernon L.  
APPLICANT: O'Mahony, Daniel J.  
APPLICANT: Lambkin, Imelda J.  
APPLICANT: Singleton, Judith  
APPLICANT: Patterson, Catherine A.  
APPLICANT: Cagney, Gerard M.  
APPLICANT: Belinka, Benjamin A.  
APPLICANT: Carter, John M.

TITLE OF INVENTION: RANDOM PEPTIDES THAT BIND TO GASTRO-

NUMBER OF SEQUENCES: 407

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSEQ Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/079,678A  
FILING DATE: 15-May-1998  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-220  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 178:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 832 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 178:  
US-09-079-678-178

Query Match 100.0%; Score 4321; DB 14; Length 832;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 832; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MILQHLHSLCLMLYLATGYQGEGKSGPLKPMFTFSIYEGQEPSQIIQFKANPVAVF 60  
Db 1 MILQHLHSLCLMLYLATGYQGEGKSGPLKPMFTFSIYEGQEPSQIIQFKANPVAVF 60

Qy 61 ELTGTDNIFVIERGLLYYRNALDRETRSTHNLQVAALDANGIIVEGVPVITIEVKDIN 120  
Db 61 ELTGTDNIFVIERGLLYYRNALDRETRSTHNLQVAALDANGIIVEGVPVITIEVKDIN 120

Qy 121 DNRPTFLQSKYEGSVQRNSRCPKPLYVNATDLDPPATPNGLYQYIYIQLPMINNVYF 180  
Db 121 DNRPTFLQSKYEGSVQRNSRCPKPLYVNATDLDPPATPNGLYQYIYIQLPMINNVYF 180

Qy 181 QINNKGTGALSITREGSQELNPAKNSYNLVISVKDMGQSGSENSFSDTTSVDIIVTENIWK 240  
Db 181 QINNKGTGALSITREGSQELNPAKNSYNLVISVKDMGQSGSENSFSDTTSVDIIVTENIWK 240

Qy 241 APKPVEMVNSTDPHPKITYVRNNDPCAQYSLVDKELPRFPFSIDQEGDIYVTOPLDR 300  
Db 241 APKPVEMVNSTDPHPKITYVRNNDPCAQYSLVDKELPRFPFSIDQEGDIYVTOPLDR 300

Qy 301 BEKDAYFYAVAKDEYKPLSYPLIEHVKKVDINDNPPTCSPVTVFEVQENERLGNISG 360  
Db 301 BEKDAYFYAVAKDEYKPLSYPLIEHVKKVDINDNPPTCSPVTVFEVQENERLGNISG 360

Qy 361 TLTAHDRDEENTANSFLNRYVEQTPKLPMDGLFLIYQYAGMLQAKOSLKKQDTPQYNL 420  
Db 361 TLTAHDRDEENTANSFLNRYVEQTPKLPMDGLFLIYQYAGMLQAKOSLKKQDTPQYNL 420

Qy 421 TIEVSDKDFKTLCFQVQINVIDINDQIPFEKSDYGNLTLAEDTNIGSTILTIQATDADEP 480  
Db 421 TIEVSDKDFKTLCFQVQINVIDINDQIPFEKSDYGNLTLAEDTNIGSTILTIQATDADEP 480

Qy 481 FTGSSKILYHIKGDSEGRGLGVDPHTNTGYVVIKKPLDFTETAASVNIKFAENPEPLV 540  
Db 481 FTGSSKILYHIKGDSEGRGLGVDPHTNTGYVVIKKPLDFTETAASVNIKFAENPEPLV 540

Qy 541 FGVYNASSFAKFTLIIVTDVNEAPQFSOHVFOAKVSEDAVIGTKVGNVYAKDPEGLDISY 600  
Db 541 FGVYNASSFAKFTLIIVTDVNEAPQFSOHVFOAKVSEDAVIGTKVGNVYAKDPEGLDISY 600

Qy 601 SLRGDTRGWLKIDHWTCGIFSVAPLDRAGSPYRVQVVATEVGGSSLSVSEFFHILIMDV 660  
Db 601 SLRGDTRGWLKIDHWTCGIFSVAPLDRAGSPYRVQVVATEVGGSSLSVSEFFHILIMDV 660

Qy 661 NDNPRLLAKDYTGFLFFCHPLSAPGSLIFEATDDOHLFRGPHFTFSLGSGSLONDWEYSK 720  
Db 661 NDNPRLLAKDYTGFLFFCHPLSAPGSLIFEATDDOHLFRGPHFTFSLGSGSLONDWEYSK 720

Qy 721 INGT HARLSRHTDPEERAYVVLIRINDGRPPLEGIVSLPVTFCSCVEGSCFRPAGHOT 780  
Db 721 INGT HARLSRHTDPEERAYVVLIRINDGRPPLEGIVSLPVTFCSCVEGSCFRPAGHOT 780

Qy 781 GIPVGMVAGILLTLLVIGIILAVVFIIRIKDKGKNVESAQAASEVKPLRS 832  
Db 781 GIPVGMVAGILLTLLVIGIILAVVFIIRIKDKGKNVESAQAASEVKPLRS 832

RESULT 5  
US-09-079-723-178  
; Sequence 178, Application US/09079723  
; GENERAL INFORMATION:  
; APPLICANT: Alvarez, Vernon L.  
; APPLICANT: O'Mahony, Daniel J.  
; APPLICANT: Lambkin, Imelda J.  
; APPLICANT: Singleton, Judith  
; APPLICANT: Patterson, Catherine A.  
; APPLICANT: Cagney, Gerard M.  
; APPLICANT: Belinka, Benjamin A.  
; APPLICANT: Carter, John M.  
; TITLE OF INVENTION: RANDOM PEPTIDES THAT BIND TO GASTRO-  
; TITLE OF INVENTION: INTESTINAL TRACT (GIT) TRANSPORT RECEPTORS AND RELATED METH  
; NUMBER OF SEQUENCES: 265  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/079,723  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mistrock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 1101-219  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 178:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 832 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; US-09-079-723-178

Query Match 100.0%; Score 4321; DB 14; Length 832;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 832; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MILOAHLHSLCLMLYLATGYGQSGKPSGLPKMTFSIYEGQEPSQIIFQKAPPPAVTF 60  
DB 1 MILOAHLHSLCLMLYLATGYGQSGKPSGLPKMTFSIYEGQEPSQIIFQKAPPPAVTF 60  
QY 61 ELTGETDNIFVIEREGLLYNRALDRETRSTHNLQVAALDANGIIVEGVPITIEVKDIN 120  
DB 61 ELTGETDNIFVIEREGLLYNRALDRETRSTHNLQVAALDANGIIVEGVPITIEVKDIN 120  
QY 121 DNRPTFLQSKYEGSVRONSPRGKPFYVNDATDLDPPATNGQLYQYIOLPMLNNMYF 180  
DB 121 DNRPTFLQSKYEGSVRONSPRGKPFYVNDATDLDPPATNGQLYQYIOLPMLNNMYF 180  
QY 181 QINNKTAISLTRGSOELNPAKNSYLVSVKDMGQSGNSFSDTTSVDIIVTENIWK 240  
DB 181 QINNKTAISLTRGSOELNPAKNSYLVSVKDMGQSGNSFSDTTSVDIIVTENIWK 240  
QY 241 APKPEVMVENSTDPHPKIKITQVRWNPDAQYSLVDKELPRFPESIDQEGDIYVTPDLR 300  
DB 241 APKPEVMVENSTDPHPKIKITQVRWNPDAQYSLVDKELPRFPESIDQEGDIYVTPDLR 300  
QY 301 EERDAYVYAVAKDEYKPLSYLPIEIHVKVDINDNPTCPSPVTVFEVQENERLGNISG 360  
DB 301 EERDAYVYAVAKDEYKPLSYLPIEIHVKVDINDNPTCPSPVTVFEVQENERLGNISG 360  
QY 361 TLTAHREDEENTANSFLNYRIVEQTPKLPMDGLFLIOTYAGMLQAKQSLKQDTPQYNL 420  
DB 361 TLTAHREDEENTANSFLNYRIVEQTPKLPMDGLFLIOTYAGMLQAKQSLKQDTPQYNL 420  
QY 421 TIEVSDKDFKTLFCFQVINVDINDQIPIFEKSDYGNLTAEADTNGISTILTITQATDAEP 480  
DB 421 TIEVSDKDFKTLFCFQVINVDINDQIPIFEKSDYGNLTAEADTNGISTILTITQATDAEP 480  
QY 481 FTGSSKILYHIKGDSEGRGLVDTPHTNGYVIKKPLDFTAAVSNIVFKAENPEPLV 540  
DB 481 FTGSSKILYHIKGDSEGRGLVDTPHTNGYVIKKPLDFTAAVSNIVFKAENPEPLV 540  
QY 541 FGKYNASSFAKFTLIYTDVNEAPQFSQHVFOAKVSEDVAIGTKVGNVTAKDPEGLDISY 600  
DB 541 FGKYNASSFAKFTLIYTDVNEAPQFSQHVFOAKVSEDVAIGTKVGNVTAKDPEGLDISY 600  
QY 601 SLRGDTRGWLKIDHVTGEISVAPLDREAGSPYRVQVAVTEVGGSSLSYSEFHLILMDV 660  
DB 601 SLRGDTRGWLKIDHVTGEISVAPLDREAGSPYRVQVAVTEVGGSSLSYSEFHLILMDV 660  
QY 661 NDNPRLAKDYTLGFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720  
DB 661 NDNPRLAKDYTLGFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720  
QY 721 INGTHARLSTRHTDFEERAYVVLIRINDGRRPPLLEGIVSLPVTFCSCVEGSCFRPAGHOT 780  
DB 721 INGTHARLSTRHTDFEERAYVVLIRINDGRRPPLLEGIVSLPVTFCSCVEGSCFRPAGHOT 780  
QY 781 GIPTVGNAVGLITLLVIGIILAVFIRIKKDKGNVESAQASEVKPLRS 832  
DB 781 GIPTVGNAVGLITLLVIGIILAVFIRIKKDKGNVESAQASEVKPLRS 832

RESULT 6  
US-09-079-723-178

; Sequence 178, Application US/09079723A  
; GENERAL INFORMATION:

; APPLICANT: Alvarez, Vernon L.

; O'Mahony, Daniel J.

; Lambkin, Imelda J.

; Singleton, Judith

; Patterson, Catherine A.

; Cagney, Gerard M.

; Belinka, Benjamin A.

; Carter, John M.

; TITLE OF INVENTION: RANDOM PEPTIDES THAT BIND TO GASTRO-

; INTESTINAL TRACT (GIT) TRANSPORT RECEPTORS AND RELATED METH

; NUMBER OF SEQUENCES: 407

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/079,723A  
; FILING DATE: 15-May-1998  
; CLASSIFICATION: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mirock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 1101-219  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 178:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 832 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: <Unknown>  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 178:  
US-09-079-723-178

Query Match 100.0%; Score 4321; DB 14; Length 832;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 832; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MILOAHLHSLCLMLYLATGYGQSGKPSGLPKMTFSIYEGQEPSQIIFQKAPPPAVTF 60  
DB 1 MILOAHLHSLCLMLYLATGYGQSGKPSGLPKMTFSIYEGQEPSQIIFQKAPPPAVTF 60  
QY 61 ELTGETDNIFVIEREGLLYNRALDRETRSTHNLQVAALDANGIIVEGVPITIEVKDIN 120  
DB 61 ELTGETDNIFVIEREGLLYNRALDRETRSTHNLQVAALDANGIIVEGVPITIEVKDIN 120  
QY 121 DNRPTFLQSKYEGSVRONSPRGKPFYVNDATDLDPPATNGQLYQYIOLPMLNNMYF 180  
DB 121 DNRPTFLQSKYEGSVRONSPRGKPFYVNDATDLDPPATNGQLYQYIOLPMLNNMYF 180  
QY 181 QINNKTAISLTRGSOELNPAKNSYLVSVKDMGQSGNSFSDTTSVDIIVTENIWK 240  
DB 181 QINNKTAISLTRGSOELNPAKNSYLVSVKDMGQSGNSFSDTTSVDIIVTENIWK 240  
QY 241 APKPEVMVENSTDPHPKIKITQVRWNPDAQYSLVDKELPRFPESIDQEGDIYVTPDLR 300  
DB 241 APKPEVMVENSTDPHPKIKITQVRWNPDAQYSLVDKELPRFPESIDQEGDIYVTPDLR 300  
QY 301 EERDAYVYAVAKDEYKPLSYLPIEIHVKVDINDNPTCPSPVTVFEVQENERLGNISG 360  
DB 301 EERDAYVYAVAKDEYKPLSYLPIEIHVKVDINDNPTCPSPVTVFEVQENERLGNISG 360  
QY 361 TLTAHREDEENTANSFLNYRIVEQTPKLPMDGLFLIOTYAGMLQAKQSLKQDTPQYNL 420  
DB 361 TLTAHREDEENTANSFLNYRIVEQTPKLPMDGLFLIOTYAGMLQAKQSLKQDTPQYNL 420  
QY 421 TIEVSDKDFKTLFCFQVINVDINDQIPIFEKSDYGNLTAEADTNGISTILTITQATDAEP 480  
DB 421 TIEVSDKDFKTLFCFQVINVDINDQIPIFEKSDYGNLTAEADTNGISTILTITQATDAEP 480  
QY 481 FTGSSKILYHIKGDSEGRGLVDTPHTNGYVIKKPLDFTAAVSNIVFKAENPEPLV 540  
DB 481 FTGSSKILYHIKGDSEGRGLVDTPHTNGYVIKKPLDFTAAVSNIVFKAENPEPLV 540

```

Qy 541 FGVKNASSFAKFTLIYTDVNEAPQFSQHVFAQKVEDVAIGTKVGNVTAKDPEGLDISY 600
Db 541 FGVKNASSFAKFTLIYTDVNEAPQFSQHVFAQKVEDVAIGTKVGNVTAKDPEGLDISY 600
Qy 601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVADEVGSSLSVSEFHLILMDV 660
Db 601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVADEVGSSLSVSEFHLILMDV 660
Qy 661 NDNPPRLAKDYTGGLFFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEYSK 720
Db 661 NDNPPRLAKDYTGGLFFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEYSK 720
Qy 721 INGTHARLSTRHTDFEERAYVVLIRINDGRRPPLLEGIVSLPVTFCSCVEGSCFRPAGHOT 780
Db 721 INGTHARLSTRHTDFEERAYVVLIRINDGRRPPLLEGIVSLPVTFCSCVEGSCFRPAGHOT 780
Qy 781 GIPVGMVAGILLTLLVIGIILAVVFIRIKKDKGNVESQAQSEVKPLRS 832
Db 781 GIPVGMVAGILLTLLVIGIILAVVFIRIKKDKGNVESQAQSEVKPLRS 832

RESULT 7
US-09-079-819-178
; Sequence 178, Application US/09079819
; GENERAL INFORMATION:
; APPLICANT: Alvarez, Vernon L.
; APPLICANT: O'Mahony, Daniel J.
; APPLICANT: Lambkin, Imelda J.
; APPLICANT: Singleton, Judith
; APPLICANT: Patterson, Catherine A.
; APPLICANT: Cagney, Gerard M.
; APPLICANT: Belinka, Benjamin A.
; APPLICANT: Carter, John M.
; TITLE OF INVENTION: RANDOM PEPTIDES THAT BIND TO GASTRO-
; TITLE OF INVENTION: INTESTINAL TRACT (GIT) TRANSPORT RECEPTORS AND RELATED METHODS
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/079,819
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-209
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 178:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 832 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-09-079-819-178

```

```

Query Match 100.0%; Score 4321; DB 14; Length 832;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 832; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 MILQAHLSLCLMLLYLATGYQGBKFGSPDKPMTFSIYEGQEPSQIIIFQKAPPAVTF 60
Db 1 MILQAHLSLCLMLLYLATGYQGBKFGSPDKPMTFSIYEGQEPSQIIIFQKAPPAVTF 60
Qy 61 ELTGEDTNIFVIEREGLLYNNRDLRETRSHNLQVAALDANGIIIEGVPVITTEVKDIN 120
Db 61 ELTGEDTNIFVIEREGLLYNNRDLRETRSHNLQVAALDANGIIIEGVPVITTEVKDIN 120
Qy 121 DNRFTFLQSKYEGSVQRNSRQSGKPLFYVNAATDLDDPATPNQGLYYQIVQLPMMNNMYF 180
Db 121 DNRFTFLQSKYEGSVQRNSRQSGKPLFYVNAATDLDDPATPNQGLYYQIVQLPMMNNMYF 180
Qy 181 QINNKTKAISLTRGSOBELNPAKNPSYNLVTSVKDMGQSQSENSFSDTTSVDIIIVTENIWK 240
Db 181 QINNKTKAISLTRGSOBELNPAKNPSYNLVTSVKDMGQSQSENSFSDTTSVDIIIVTENIWK 240
Qy 241 APKPVEMVNSTDPHPKIKITQVRNNDPGAQSLVDKELPRPFPSIDQEGDIYVTPQDNR 300
Db 241 APKPVEMVNSTDPHPKIKITQVRNNDPGAQSLVDKELPRPFPSIDQEGDIYVTPQDNR 300
Qy 301 EEKDAYVYAVAKDEYGGKPLSYPLEIHVKVDINDNPPTCPSVTVFEVQENERLGNISG 360
Db 301 EEKDAYVYAVAKDEYGGKPLSYPLEIHVKVDINDNPPTCPSVTVFEVQENERLGNISG 360
Qy 361 TLTAHDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYYAGMLQAKOSLKKQDTPQYNL 420
Db 361 TLTAHDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYYAGMLQAKOSLKKQDTPQYNL 420
Qy 421 TIEVSDKDFKTLCFVQINVIDINQIPIFEKSDYGNLTAEEDTNGISTILTIQATDADEP 480
Db 421 TIEVSDKDFKTLCFVQINVIDINQIPIFEKSDYGNLTAEEDTNGISTILTIQATDADEP 480
Qy 481 FTGSKILYHIKDGSEGRGLGVDTPHTNTGYVVIKKPLDFETAASVNIIVFKAENPEPLV 540
Db 481 FTGSKILYHIKDGSEGRGLGVDTPHTNTGYVVIKKPLDFETAASVNIIVFKAENPEPLV 540
Qy 541 FGVKNASSFAKFTLIYTDVNEAPQFSQHVFAQKVEDVAIGTKVGNVTAKDPEGLDISY 600
Db 541 FGVKNASSFAKFTLIYTDVNEAPQFSQHVFAQKVEDVAIGTKVGNVTAKDPEGLDISY 600
Qy 601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVADEVGSSLSVSEFHLILMDV 660
Db 601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVADEVGSSLSVSEFHLILMDV 660
Qy 661 NDNPPRLAKDYTGGLFFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEYSK 720
Db 661 NDNPPRLAKDYTGGLFFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEYSK 720
Qy 721 INGTHARLSTRHTDFEERAYVVLIRINDGRRPPLLEGIVSLPVTFCSCVEGSCFRPAGHOT 780
Db 721 INGTHARLSTRHTDFEERAYVVLIRINDGRRPPLLEGIVSLPVTFCSCVEGSCFRPAGHOT 780
Qy 781 GIPVGMVAGILLTLLVIGIILAVVFIRIKKDKGNVESQAQSEVKPLRS 832
Db 781 GIPVGMVAGILLTLLVIGIILAVVFIRIKKDKGNVESQAQSEVKPLRS 832

RESULT 8
US-09-079-819-178
; Sequence 178, Application US/09079819
; GENERAL INFORMATION:
; APPLICANT: Alvarez, Vernon L.
; APPLICANT: O'Mahony, Daniel J.
; APPLICANT: Lambkin, Imelda J.
; APPLICANT: Singleton, Judith
; APPLICANT: Patterson, Catherine A.
; APPLICANT: Cagney, Gerard M.
; APPLICANT: Belinka, Benjamin A.
; APPLICANT: Carter, John M.
; TITLE OF INVENTION: RANDOM PEPTIDES THAT BIND TO GASTRO-
; TITLE OF INVENTION: INTESTINAL TRACT (GIT) TRANSPORT RECEPTORS AND RELATED M
; NUMBER OF SEQUENCES: 407
; CORRESPONDENCE ADDRESS:

```



```
Db 361 TLTAHDEENTANSLNRYVEOTPKLPMDGLFLIOTYAGMLQAKOSLKKQDTPQYNL 420
QY 421 TIEVSDRDKFTLCFVQINVDINDQIPIFEKSDYGNLTLAEDTNGTILTIQATDADEP 480
Db 421 TIEVSDRDKFTLCFVQINVDINDQIPIFEKSDYGNLTLAEDTNGTILTIQATDADEP 480
QY 481 FTGSKILYHIKDGSEGRGLGVDTPHTNTGYVVIKKPLDFETAASNIVFKAENPEPLV 540
Db 481 FTGSKILYHIKDGSEGRGLGVDTPHTNTGYVVIKKPLDFETAASNIVFKAENPEPLV 540
QY 541 FGKYNASSFAKFTLIIVTDVNEAPQFSOHVFOAKVSEDAVAICTKGVNVTAKDPPEGLDISY 600
Db 541 FGKYNASSFAKFTLIIVTDVNEAPQFSOHVFOAKVSEDAVAICTKGVNVTAKDPPEGLDISY 600
QY 601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVVATEVGGSSLSVSEFHLILMDV 660
Db 601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVVATEVGGSSLSVSEFHLILMDV 660
QY 661 NDNPRLAKDYGTLFFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720
Db 661 NDNPRLAKDYGTLFFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720
QY 721 INGTARLSTRHTDFEERAYVVLIRINDGRRPPLLEGIVSLPVTFCSCVEGSCFRPAGHOT 780
Db 721 INGTARLSTRHTDFEERAYVVLIRINDGRRPPLLEGIVSLPVTFCSCVEGSCFRPAGHOT 780
QY 781 GIPTVGMAGVILLTLLVIGIILAVVFIIRIKKDKGNVESQAQSEVKPLRS 832
Db 781 GIPTVGMAGVILLTLLVIGIILAVVFIIRIKKDKGNVESQAQSEVKPLRS 832

RESULT 10
US-09-649-811-1081
; Sequence 1081, Application US/09649811
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C11
; CURRENT APPLICATION NUMBER: US/09/649,811
; CURRENT FILING DATE: 2000-08-28
; NUMBER OF SEQ ID NOS: 1083
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1081
; LENGTH: 832
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-649-811-1081

Query Match 100.0%; Score 4321; DB 20; Length 832;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 832; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MILQAHLSLCLLMYLATGYGQEGKFSGLPKPMTFSIYEGQEPSQIIFQKPNPAAVTF 60
Db 1 MILQAHLSLCLLMYLATGYGQEGKFSGLPKPMTFSIYEGQEPSQIIFQKPNPAAVTF 60

QY 61 ELTGETDNIFVIEREGLLYNRRALDRETRSHNLQVAALDANGIIVEGPPITIEVKDIN 120
Db 61 ELTGETDNIFVIEREGLLYNRRALDRETRSHNLQVAALDANGIIVEGPPITIEVKDIN 120

QY 121 DNRPTFLOSQYEGSVRONSRPGKFPFLYNATDLDPTATPNQGLYQYIVQLPMTNNWYF 180
Db 121 DNRPTFLOSQYEGSVRONSRPGKFPFLYNATDLDPTATPNQGLYQYIVQLPMTNNWYF 180

QY 181 QINNKTCALISLTRGSOELNPAKNPSNLYISVKDMGQSENFSDDTTSVDIIVTENIWK 240
Db 181 QINNKTCALISLTRGSOELNPAKNPSNLYISVKDMGQSENFSDDTTSVDIIVTENIWK 240
```

```
Db 181 QINNKTCALISLTRGSOELNPAKNPSNLYISVKDMGQSENFSDDTTSVDIIVTENIWK 240
QY 241 APKEVENVSTDPHPKIKITQVRWNPFGAQYSLYVDKELKPLRPFPSIDQEGDIYVTQPLDR 300
Db 241 APKEVENVSTDPHPKIKITQVRWNPFGAQYSLYVDKELKPLRPFPSIDQEGDIYVTQPLDR 300
QY 301 EEKDAYVYFAVAKDEYKPLSYPLEIHHVKYKDIINDNPPTCSPVTVFEVQNERLGNISIG 360
Db 301 EEKDAYVYFAVAKDEYKPLSYPLEIHHVKYKDIINDNPPTCSPVTVFEVQNERLGNISIG 360
QY 361 TLTAHDEENTANSLNRYVEOTPKLPMDGLFLIOTYAGMLQAKOSLKKQDTPQYNL 420
Db 361 TLTAHDEENTANSLNRYVEOTPKLPMDGLFLIOTYAGMLQAKOSLKKQDTPQYNL 420
QY 421 TIEVSDRDKFTLCFVQINVDINDQIPIFEKSDYGNLTLAEDTNGTILTIQATDADEP 480
Db 421 TIEVSDRDKFTLCFVQINVDINDQIPIFEKSDYGNLTLAEDTNGTILTIQATDADEP 480
QY 481 FTGSKILYHIKDGSEGRGLGVDTPHTNTGYVVIKKPLDFETAASNIVFKAENPEPLV 540
Db 481 FTGSKILYHIKDGSEGRGLGVDTPHTNTGYVVIKKPLDFETAASNIVFKAENPEPLV 540
QY 541 FGKYNASSFAKFTLIIVTDVNEAPQFSOHVFOAKVSEDAVAICTKGVNVTAKDPPEGLDISY 600
Db 541 FGKYNASSFAKFTLIIVTDVNEAPQFSOHVFOAKVSEDAVAICTKGVNVTAKDPPEGLDISY 600
QY 601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVVATEVGGSSLSVSEFHLILMDV 660
Db 601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVVATEVGGSSLSVSEFHLILMDV 660
QY 661 NDNPRLAKDYGTLFFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720
Db 661 NDNPRLAKDYGTLFFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720
QY 721 INGTARLSTRHTDFEERAYVVLIRINDGRRPPLLEGIVSLPVTFCSCVEGSCFRPAGHOT 780
Db 721 INGTARLSTRHTDFEERAYVVLIRINDGRRPPLLEGIVSLPVTFCSCVEGSCFRPAGHOT 780
QY 781 GIPTVGMAGVILLTLLVIGIILAVVFIIRIKKDKGNVESQAQSEVKPLRS 832
Db 781 GIPTVGMAGVILLTLLVIGIILAVVFIIRIKKDKGNVESQAQSEVKPLRS 832

RESULT 11
US-09-791-537-48350
; Sequence 48350, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 48350
; LENGTH: 832
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-791-537-48350

Query Match 100.0%; Score 4321; DB 21; Length 832;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 832; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MILQAHLSLCLLMYLATGYGQEGKFSGLPKPMTFSIYEGQEPSQIIFQKPNPAAVTF 60
Db 1 MILQAHLSLCLLMYLATGYGQEGKFSGLPKPMTFSIYEGQEPSQIIFQKPNPAAVTF 60

QY 61 ELTGETDNIFVIEREGLLYNRRALDRETRSHNLQVAALDANGIIVEGPPITIEVKDIN 120
Db 61 ELTGETDNIFVIEREGLLYNRRALDRETRSHNLQVAALDANGIIVEGPPITIEVKDIN 120
```

Db 61 ELTGETDNIFVIEREGLLYNNRDLRETRSTHNLQVAALDANGIIIEGVPPIITIEVKDIN 120  
QY 121 DNRPTFLQSKYEGSVRONSRPGKPFLLYVYNATDLDPATNGQLYQIYQIOLPMNNMYF 180  
Db 121 DNRPTFLQSKYEGSVRONSRPGKPFLLYVYNATDLDPATNGQLYQIYQIOLPMNNMYF 180  
QY 181 QINNKTAISLTREGSOELNPAKNPSYLVSKDMGGQSENFSFSDTTSVDIIIVTENIWK 240  
Db 181 QINNKTAISLTREGSOELNPAKNPSYLVSKDMGGQSENFSFSDTTSVDIIIVTENIWK 240  
QY 241 APKPVEMVENSTDPHPKIKITQVRWNPFGAQYSLVDKEKLPFRFPFSDIOEGDIYVTPQLDR 300  
Db 241 APKPVEMVENSTDPHPKIKITQVRWNPFGAQYSLVDKEKLPFRFPFSDIOEGDIYVTPQLDR 300  
QY 301 EERDAYVYAVAKDEYKPLSYPLEIHVKVDINDNPPCTCPSPTVTFEVOENERLGNISG 360  
Db 301 EERDAYVYAVAKDEYKPLSYPLEIHVKVDINDNPPCTCPSPTVTFEVOENERLGNISG 360  
QY 361 TLTAHOREDEENTANSFLNYRIVEQTPKLPMDGLFLIQTAYAGMLQAKOSLKQDTPQYNL 420  
Db 361 TLTAHOREDEENTANSFLNYRIVEQTPKLPMDGLFLIQTAYAGMLQAKOSLKQDTPQYNL 420  
QY 421 TIEVSKDKFTLCFVQINVIDINDQIPIFEKSDYGNLTAEADTNIIGSTILTIQATDADEP 480  
Db 421 TIEVSKDKFTLCFVQINVIDINDQIPIFEKSDYGNLTAEADTNIIGSTILTIQATDADEP 480  
QY 481 FTGSSKILYHIKGDSEGRGLVDTDPHTNTGYVYIKKPLDFETAASVNIIVFKAENPEPLV 540  
Db 481 FTGSSKILYHIKGDSEGRGLVDTDPHTNTGYVYIKKPLDFETAASVNIIVFKAENPEPLV 540  
QY 541 FGKYNASSFAKFTLIVTDVNEAPQFSQHVFOAKVSEDVAIGTKVGNVTAKDPGLDISY 600  
Db 541 FGKYNASSFAKFTLIVTDVNEAPQFSQHVFOAKVSEDVAIGTKVGNVTAKDPGLDISY 600  
QY 601 SLRGDTRGWLKIDHVTGEIFSVAPLDRAGSPYRVQVAVTEVGGSSLSVSEFHLILMDV 660  
Db 601 SLRGDTRGWLKIDHVTGEIFSVAPLDRAGSPYRVQVAVTEVGGSSLSVSEFHLILMDV 660  
QY 661 NDPPRLAKDYTGLEFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720  
Db 661 NDPPRLAKDYTGLEFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720  
QY 721 INGTARLSRHTDFFERAYVWLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHOT 780  
Db 721 INGTARLSRHTDFFERAYVWLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHOT 780  
QY 781 GIPTVGMAGVILLTLLVIGIILAVVFIKDKGKDNVESQAASEVKPLRS 832  
Db 781 GIPTVGMAGVILLTLLVIGIILAVVFIKDKGKDNVESQAASEVKPLRS 832

RESULT 12  
US-09-833-263-1081  
; Sequence 1081, Application US/09833263  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Aijun  
; APPLICANT: Clapper, Jonathan D.  
; APPLICANT: Stolk, John A.  
; APPLICANT: Meagher, Madeleine J.  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND  
; FILE REFERENCE: 210121.471C12  
; CURRENT APPLICATION NUMBER: US/09/833,263  
; CURRENT FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 1093  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1081  
; LENGTH: 832  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-833-263-1081

Query Match 100.0%; Score 4321; DB 22; Length 832;

Best Local Similarity 100.0%; Pred. No. 0;  
Matches 832; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MILOAHLHSCLLLMLYLATGYGQEGKESGPKLPMPTFSIYEGQEPSQIIFOKANPPAVTF 60  
Db 1 MILOAHLHSCLLLMLYLATGYGQEGKESGPKLPMPTFSIYEGQEPSQIIFOKANPPAVTF 60  
QY 61 ELTGETDNIFVIEREGLLYNNRDLRETRSTHNLQVAALDANGIIIEGVPPIITIEVKDIN 120  
Db 61 ELTGETDNIFVIEREGLLYNNRDLRETRSTHNLQVAALDANGIIIEGVPPIITIEVKDIN 120  
QY 121 DNRPTFLQSKYEGSVRONSRPGKPFLLYVYNATDLDPATNGQLYQIYQIOLPMNNMYF 180  
Db 121 DNRPTFLQSKYEGSVRONSRPGKPFLLYVYNATDLDPATNGQLYQIYQIOLPMNNMYF 180  
QY 181 QINNKTAISLTREGSOELNPAKNPSYLVSKDMGGQSENFSFSDTTSVDIIIVTENIWK 240  
Db 181 QINNKTAISLTREGSOELNPAKNPSYLVSKDMGGQSENFSFSDTTSVDIIIVTENIWK 240  
QY 241 APKPVEMVENSTDPHPKIKITQVRWNPFGAQYSLVDKEKLPFRFPFSDIOEGDIYVTPQLDR 300  
Db 241 APKPVEMVENSTDPHPKIKITQVRWNPFGAQYSLVDKEKLPFRFPFSDIOEGDIYVTPQLDR 300  
QY 301 EERDAYVYAVAKDEYKPLSYPLEIHVKVDINDNPPCTCPSPTVTFEVOENERLGNISG 360  
Db 301 EERDAYVYAVAKDEYKPLSYPLEIHVKVDINDNPPCTCPSPTVTFEVOENERLGNISG 360  
QY 361 TLTAHOREDEENTANSFLNYRIVEQTPKLPMDGLFLIQTAYAGMLQAKOSLKQDTPQYNL 420  
Db 361 TLTAHOREDEENTANSFLNYRIVEQTPKLPMDGLFLIQTAYAGMLQAKOSLKQDTPQYNL 420  
QY 421 TIEVSKDKFTLCFVQINVIDINDQIPIFEKSDYGNLTAEADTNIIGSTILTIQATDADEP 480  
Db 421 TIEVSKDKFTLCFVQINVIDINDQIPIFEKSDYGNLTAEADTNIIGSTILTIQATDADEP 480  
QY 481 FTGSSKILYHIKGDSEGRGLVDTDPHTNTGYVYIKKPLDFETAASVNIIVFKAENPEPLV 540  
Db 481 FTGSSKILYHIKGDSEGRGLVDTDPHTNTGYVYIKKPLDFETAASVNIIVFKAENPEPLV 540  
QY 541 FGKYNASSFAKFTLIVTDVNEAPQFSQHVFOAKVSEDVAIGTKVGNVTAKDPGLDISY 600  
Db 541 FGKYNASSFAKFTLIVTDVNEAPQFSQHVFOAKVSEDVAIGTKVGNVTAKDPGLDISY 600  
QY 601 SLRGDTRGWLKIDHVTGEIFSVAPLDRAGSPYRVQVAVTEVGGSSLSVSEFHLILMDV 660  
Db 601 SLRGDTRGWLKIDHVTGEIFSVAPLDRAGSPYRVQVAVTEVGGSSLSVSEFHLILMDV 660  
QY 661 NDPPRLAKDYTGLEFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720  
Db 661 NDPPRLAKDYTGLEFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720  
QY 721 INGTARLSRHTDFFERAYVWLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHOT 780  
Db 721 INGTARLSRHTDFFERAYVWLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHOT 780  
QY 781 GIPTVGMAGVILLTLLVIGIILAVVFIKDKGKDNVESQAASEVKPLRS 832  
Db 781 GIPTVGMAGVILLTLLVIGIILAVVFIKDKGKDNVESQAASEVKPLRS 832

RESULT 13  
US-09-922-217-1081  
; Sequence 1081, Application US/09922217  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Meagher, Madeleine Joy  
; APPLICANT: Stolk, John A.  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Jiang, Yudi  
; APPLICANT: Smith, Carole Lynn

```

; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922,217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1081
; LENGTH: 832
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-922-217-1081

Query Match 100.0%; Score 4321; DB 23; Length 832;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 832; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MILQAHLSLCLMLYLATGYGEGKESGPKPMTFSIYEGQEPSQIIFOKANPPAVTF 60
Db 1 MILQAHLSLCLMLYLATGYGEGKESGPKPMTFSIYEGQEPSQIIFOKANPPAVTF 60
QY 61 ELTGTONIFVIEREGLLYNNRDLDRSTRTHNLQVAALDANGIIIVGPPVITIEVKDIN 120
Db 61 ELTGTONIFVIEREGLLYNNRDLDRSTRTHNLQVAALDANGIIIVGPPVITIEVKDIN 120
QY 121 DNRPTFLQSYEGSVQRNSRPGKPLVYNATDLDDPATPNQOLYQIVQIOLPMINNNMYF 180
Db 121 DNRPTFLQSYEGSVQRNSRPGKPLVYNATDLDDPATPNQOLYQIVQIOLPMINNNMYF 180
QY 121 DNRPTFLQSYEGSVQRNSRPGKPLVYNATDLDDPATPNQOLYQIVQIOLPMINNNMYF 180
Db 121 DNRPTFLQSYEGSVQRNSRPGKPLVYNATDLDDPATPNQOLYQIVQIOLPMINNNMYF 180
QY 181 QINNKTCAISLTRGSOQLNPAKNPSNLVSVKDMGQSENSESDTTSVDIIIVTENIWK 240
Db 181 QINNKTCAISLTRGSOQLNPAKNPSNLVSVKDMGQSENSESDTTSVDIIIVTENIWK 240
QY 241 APKPEVMVENSTDPHPKIKITQVRWNDPGAQYSLVDKKEKLPFPFSDIOEGDIYVTPQLDR 300
Db 241 APKPEVMVENSTDPHPKIKITQVRWNDPGAQYSLVDKKEKLPFPFSDIOEGDIYVTPQLDR 300
QY 301 EEKDAYVYFAVAKDEYKPLSYPLEIHVKVDINDNPTCPSPVTVFEVQENERLGNISG 360
Db 301 EEKDAYVYFAVAKDEYKPLSYPLEIHVKVDINDNPTCPSPVTVFEVQENERLGNISG 360
QY 361 TLTAHDEENTANSFLNYRIVEQTPKLPMDGLFLIQTAYAGMLQAKOSLKKQDTPQYNL 420
Db 361 TLTAHDEENTANSFLNYRIVEQTPKLPMDGLFLIQTAYAGMLQAKOSLKKQDTPQYNL 420
QY 421 TIEVSDKDFKTLCFVQINVDINDQIPIFEKSDYGNLTLAEDTNGISTILTIOATDADEP 480
Db 421 TIEVSDKDFKTLCFVQINVDINDQIPIFEKSDYGNLTLAEDTNGISTILTIOATDADEP 480
QY 481 FTGSSKILYHIKGDSEGRGVDPDHTNTGYVITIKKPLDFETAASNIVFKAENPEPLV 540
Db 481 FTGSSKILYHIKGDSEGRGVDPDHTNTGYVITIKKPLDFETAASNIVFKAENPEPLV 540
QY 541 FGKYNASSFAKFTLIIVTDVNEAPQFSQHVFAKVSDEVAIGTKVGNVTAKDPGLDISY 600
Db 541 FGKYNASSFAKFTLIIVTDVNEAPQFSQHVFAKVSDEVAIGTKVGNVTAKDPGLDISY 600
QY 601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSVSEFHLILMDV 660
Db 601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSVSEFHLILMDV 660
QY 661 NDNPRKADYTGFLFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720
Db 661 NDNPRKADYTGFLFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720
QY 721 INGTARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTSCVCEGSCFRPAGHOT 780
Db 721 INGTARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTSCVCEGSCFRPAGHOT 780
QY 781 GIPTVGMVAGILLTLLVIGIILAVFIRIKKDKGKNVESQAQASEVKPLRS 832
```

```

Db 781 GIPTVGMVAGILLTLLVIGIILAVFIRIKKDKGKNVESQAQASEVKPLRS 832
RESULT 14
US-09-443-780C-15
; Sequence 15, Application US/09443780C
; GENERAL INFORMATION:
; APPLICANT: O'Mahony, Daniel J
; APPLICANT: Seveso, Michela
; TITLE OF INVENTION: Antibodies To Peptides That Target GIT Receptors And Related M
; FILE REFERENCE: E1067/20037
; CURRENT APPLICATION NUMBER: US/09/443,780C
; CURRENT FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: US 60/109,036
; PRIOR FILING DATE: 1998-11-19
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 832
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HPT1 receptor
; US-09-443-780C-15

Query Match 100.0%; Score 4321; DB 23; Length 832;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 832; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MILQAHLSLCLMLYLATGYGEGKESGPKPMTFSIYEGQEPSQIIFOKANPPAVTF 60
Db 1 MILQAHLSLCLMLYLATGYGEGKESGPKPMTFSIYEGQEPSQIIFOKANPPAVTF 60
QY 61 ELTGTONIFVIEREGLLYNNRDLDRSTRTHNLQVAALDANGIIIVGPPVITIEVKDIN 120
Db 61 ELTGTONIFVIEREGLLYNNRDLDRSTRTHNLQVAALDANGIIIVGPPVITIEVKDIN 120
QY 121 DNRPTFLQSYEGSVQRNSRPGKPLVYNATDLDDPATPNQOLYQIVQIOLPMINNNMYF 180
Db 121 DNRPTFLQSYEGSVQRNSRPGKPLVYNATDLDDPATPNQOLYQIVQIOLPMINNNMYF 180
QY 181 QINNKTCAISLTRGSOQLNPAKNPSNLVSVKDMGQSENSESDTTSVDIIIVTENIWK 240
Db 181 QINNKTCAISLTRGSOQLNPAKNPSNLVSVKDMGQSENSESDTTSVDIIIVTENIWK 240
QY 241 APKPEVMVENSTDPHPKIKITQVRWNDPGAQYSLVDKKEKLPFPFSDIOEGDIYVTPQLDR 300
Db 241 APKPEVMVENSTDPHPKIKITQVRWNDPGAQYSLVDKKEKLPFPFSDIOEGDIYVTPQLDR 300
QY 301 EEKDAYVYFAVAKDEYKPLSYPLEIHVKVDINDNPTCPSPVTVFEVQENERLGNISG 360
Db 301 EEKDAYVYFAVAKDEYKPLSYPLEIHVKVDINDNPTCPSPVTVFEVQENERLGNISG 360
QY 361 TLTAHDEENTANSFLNYRIVEQTPKLPMDGLFLIQTAYAGMLQAKOSLKKQDTPQYNL 420
Db 361 TLTAHDEENTANSFLNYRIVEQTPKLPMDGLFLIQTAYAGMLQAKOSLKKQDTPQYNL 420
QY 421 TIEVSDKDFKTLCFVQINVDINDQIPIFEKSDYGNLTLAEDTNGISTILTIOATDADEP 480
Db 421 TIEVSDKDFKTLCFVQINVDINDQIPIFEKSDYGNLTLAEDTNGISTILTIOATDADEP 480
QY 481 FTGSSKILYHIKGDSEGRGVDPDHTNTGYVITIKKPLDFETAASNIVFKAENPEPLV 540
Db 481 FTGSSKILYHIKGDSEGRGVDPDHTNTGYVITIKKPLDFETAASNIVFKAENPEPLV 540
QY 541 FGKYNASSFAKFTLIIVTDVNEAPQFSQHVFAKVSDEVAIGTKVGNVTAKDPGLDISY 600
Db 541 FGKYNASSFAKFTLIIVTDVNEAPQFSQHVFAKVSDEVAIGTKVGNVTAKDPGLDISY 600
QY 601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSVSEFHLILMDV 660
Db 601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSVSEFHLILMDV 660
```



```
QY 661 NDPPRLAKDYTGFFCHPLSAPGSLIFEATDDQHLEFRGPHFTFSLGSGSLQNDWEVSK 720
|||||
Db 661 NDPPRLAKDYTGFFCHPLSAPGSLIFEATDDQHLEFRGPHFTFSLGSGSLQNDWEVSK 720
|||||
QY 721 INGTARLSRHTDFFERAYVWLIRINDGGRPLEGIVSLPVTFCSCVEGSCFRPAGHOT 780
|||||
Db 721 INGTARLSRHTDFFERAYVWLIRINDGGRPLEGIVSLPVTFCSCVEGSCFRPAGHOT 780
|||||
QY 781 GIPTVMAVGILLTLLVIGIILAVVFIIRIKKDKGDNVESQAASEVKPLRS 832
|||||
Db 781 GIPTVMAVGILLTLLVIGIILAVVFIIRIKKDKGDNVESQAASEVKPLRS 832
|||||
```

RESULT 15

```
US-10-025-380-1081
; Sequence 1081, Application US/10025380
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yugu
; APPLICANT: Smith, Carole L.
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Skelky, Yasir A. W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick Thomas S.
; APPLICANT: Carter, Darrick
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C14
; CURRENT FILING DATE: 2001-12-19
; NUMBER OF SEQ ID NOS: 1129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1081
; LENGTH: 832
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-025-380-1081
```

```
Query Match 100.0%; Score 4321; DB 24; Length 832;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 832; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MILQAHLHSLCLMLYLATYGOEGKFGPLKPMFTSIYEGQEPSQIIFOKANPPAVTF 60
|||||
Db 1 MILQAHLHSLCLMLYLATYGOEGKFGPLKPMFTSIYEGQEPSQIIFOKANPPAVTF 60
|||||
QY 61 ELTGTDNIEVIEREGLLYNRALDRETRTHNLQVAALDANGIIEVGPPITIEVKDIN 120
|||||
Db 61 ELTGTDNIEVIEREGLLYNRALDRETRTHNLQVAALDANGIIEVGPPITIEVKDIN 120
|||||
QY 121 DNRPTFLQSKYEGSVQRNSRPGPFYVYNATDLDPATNGQLYYQIVIOLPMINNMYF 180
|||||
Db 121 DNRPTFLQSKYEGSVQRNSRPGPFYVYNATDLDPATNGQLYYQIVIOLPMINNMYF 180
|||||
QY 181 QINKKTGAISLTREGSOELNPAKNPSNLVSVKMGQSGNSFSDTTSVDIIVTENIWK 240
|||||
Db 181 QINKKTGAISLTREGSOELNPAKNPSNLVSVKMGQSGNSFSDTTSVDIIVTENIWK 240
|||||
QY 241 APKPVMVENSTDPHPKIKITQVRWNPFGAQYSLVDKEKLPFRFPFSDQEGDIYVTOPLDR 300
|||||
Db 241 APKPVMVENSTDPHPKIKITQVRWNPFGAQYSLVDKEKLPFRFPFSDQEGDIYVTOPLDR 300
|||||
QY 301 EEKDAYVYFAVAKDEYKPLSYPLEIHVKVDINDNPPTCSPVTVFEVQENRGLNSIG 360
|||||
```

```
Db 301 EEKDAYVYFAVAKDEYKPLSYPLEIHVKVDINDNPPTCSPVTVFEVQENRGLNSIG 360
QY 361 TLTAHDRDEENTANSFLNYRIVEOTPKLPMDGLFLIOTYAGMLQAKQSLKKQDTPQYNL 420
|||||
Db 361 TLTAHDRDEENTANSFLNYRIVEOTPKLPMDGLFLIOTYAGMLQAKQSLKKQDTPQYNL 420
|||||
QY 421 TIEVSDKDFKTLCFVQINVIDIQIPFEKSDYGNLTLAEDTNIGSTILTITQATDADEP 480
|||||
Db 421 TIEVSDKDFKTLCFVQINVIDIQIPFEKSDYGNLTLAEDTNIGSTILTITQATDADEP 480
|||||
QY 481 FTGSSKILYHIKGDSEGRGLGVDTPHTNTGYVLIKKPLDFETAAVSNIVFKAENPEPLV 540
|||||
Db 481 FTGSSKILYHIKGDSEGRGLGVDTPHTNTGYVLIKKPLDFETAAVSNIVFKAENPEPLV 540
|||||
QY 541 FGVYNASSFAKFTLIIVTDVNEAPQFSQHVFOAKVSEDVAIGTKVGNVTAKDPEGLDISY 600
|||||
Db 541 FGVYNASSFAKFTLIIVTDVNEAPQFSQHVFOAKVSEDVAIGTKVGNVTAKDPEGLDISY 600
|||||
QY 601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSVSEFHLILMDV 660
|||||
Db 601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSVSEFHLILMDV 660
|||||
QY 661 NDPPRLAKDYTGFFCHPLSAPGSLIFEATDDQHLEFRGPHFTFSLGSGSLQNDWEVSK 720
|||||
Db 661 NDPPRLAKDYTGFFCHPLSAPGSLIFEATDDQHLEFRGPHFTFSLGSGSLQNDWEVSK 720
|||||
QY 721 INGTARLSRHTDFFERAYVWLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHOT 780
|||||
Db 721 INGTARLSRHTDFFERAYVWLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHOT 780
|||||
QY 781 GIPTVMAVGILLTLLVIGIILAVVFIIRIKKDKGDNVESQAASEVKPLRS 832
|||||
Db 781 GIPTVMAVGILLTLLVIGIILAVVFIIRIKKDKGDNVESQAASEVKPLRS 832
|||||
```

Search completed: April 2, 2003, 16:03:05  
Job time : 440.895 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 2, 2003, 15:52:54 ; Search time 91.1781 seconds  
(without alignments)  
1318.206 Million cell updates/sec

Title: US-09-079-678-178  
Perfect score: 4321  
Sequence: 1 MILQAHLSLCLMLYLATG.....DRGKDNVESQAQSEVKPLRS 832

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 656186 seqs, 144460914 residues

Total number of hits satisfying chosen parameters: 656186

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_New.\*  
1: /cgn2\_5/ptodata/2/paa/PCT\_NEW\_COMB.pep.\*  
2: /cgn2\_5/ptodata/2/paa/US06\_NEW\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/paa/US07\_NEW\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/paa/US08\_NEW\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep.\*  
7: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	4321	100.0	832	1	PCT-US02-11475A-1081
2	4208	97.4	958	1	PCT-US02-11475A-1087
3	922.5	21.3	841	5	US-09-949-016-9669
4	908.5	21.0	807	6	US-10-230-437-188
5	908.5	21.0	807	6	US-10-063-580-98
6	908.5	21.0	807	6	US-10-219-076-188
7	908.5	21.0	807	6	US-10-219-473-188
8	908.5	21.0	807	6	US-10-015-610A-229
9	908.5	21.0	807	6	US-10-226-254A-229
10	908.5	21.0	807	6	US-10-017-253A-229
11	908.5	21.0	807	6	US-10-063-557-98
12	908.5	21.0	807	6	US-09-724-676-54141
13	743.5	17.2	664	5	US-09-724-676A-54141
14	743.5	17.2	664	5	US-09-724-676A-54141
15	699.5	16.2	786	7	US-60-452-680-19984
16	699.5	16.2	786	7	US-60-452-680-19985
17	699.5	16.2	847	5	US-09-724-676-54792
18	699.5	16.2	847	5	US-09-724-676A-54792
19	699.5	16.2	885	5	US-09-724-676-54791
20	699.5	16.2	885	5	US-09-724-676A-54791
21	699.5	16.2	906	1	PCT-US02-29560-300
22	699.5	16.2	906	1	PCT-US02-19669A-43
23	699.5	16.2	906	5	US-09-724-676-54790
24	699.5	16.2	906	5	US-09-724-676A-54790
25	699.5	16.2	906	6	US-10-245-882-300
26	694.5	16.1	735	7	US-60-452-680-19986

ALIGNMENTS

RESULT 1

PCT-US02-11475A-1081  
; Sequence 1081, Application PC/TUS0211475A  
; GENERAL INFORMATION:  
; APPLICANT: Corixa Corporation  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Meagher, Madeleine Joy  
; APPLICANT: Stolk, John A.  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Jiang, Yuqiu  
; APPLICANT: Smith, Carole L.  
; APPLICANT: King, Gordon E.  
; APPLICANT: Wang, Aijun  
; APPLICANT: Clapper, Jonathan D.  
; APPLICANT: Skeiky, Yasir A. W.  
; APPLICANT: Fanger, Gary R.  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Carter, Darrick  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS  
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE  
; FILE REFERENCE: 210121.47102PC  
; CURRENT APPLICATION NUMBER: PCT/US02/11475A  
; CURRENT FILING DATE: 2002-04-09  
; NUMBER OF SEQ ID NOS: 1129  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1081  
; LENGTH: 832  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
PCT-US02-11475A-1081

Query Match 100.0%; Score 4321; DB 1; Length 832;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 832; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MILQAHLSLCLMLYLATGVGQSGKFSGLPKMPTFSIYEGQEPSQIFQKAPPAVTF	60
Db	1	MILQAHLSLCLMLYLATGVGQSGKFSGLPKMPTFSIYEGQEPSQIFQKAPPAVTF	60
Qy	61	ELTGETDNIFVIERGLLYNRALDRETRSTHNLQVAALDANGIIVGPPVITIEVKDIN	120
Db	61	ELTGETDNIFVIERGLLYNRALDRETRSTHNLQVAALDANGIIVGPPVITIEVKDIN	120
Qy	121	DNRPFTFQSKYEGSVRNSRPGKPFLLYVNATDLDPATPNQLYQIVQLPMNNMYF	180
Db	121	DNRPFTFQSKYEGSVRNSRPGKPFLLYVNATDLDPATPNQLYQIVQLPMNNMYF	180

Qy	181	QINNKTCAISLTRGSGOELNPAKNPSYNLVISVKDMGGOSENSFSDTT	SVDIIVTENIWK	240
Db	181	QINNKTCAISLTRGSGOELNPAKNPSYNLVISVKDMGGOSENSFSDTT	SVDIIVTENIWK	240
Qy	241	APKPEVMENSTDPHPKIKTQVRWMDPGAGYSLVDKEKILPRPPFS	IDQEGDIYVTOPLDR	300
Db	241	APKPEVMENSTDPHPKIKTQVRWMDPGAGYSLVDKEKILPRPPFS	IDQEGDIYVTOPLDR	300
Qy	301	EEKDAYFYAVAKDEYCKPLSYPLETHVKVKDINDNPPTCPSPVT	FVEQNERLGNSTG	360
Db	301	EEKDAYFYAVAKDEYCKPLSYPLETHVKVKDINDNPPTCPSPVT	FVEQNERLGNSTG	360
Qy	361	TLTAHDRDEENTANFSLNYRIVQTPKLPMDGLFLIQTAGMLQAKQSL	KKQDTPQYNL	420
Db	361	TLTAHDRDEENTANFSLNYRIVEQTPKLPMDGLFLIQTAGMLQAKQSL	KKQDTPQYNL	420
Qy	421	TIEVSDKDFKTLCFVQINVIDINDQIPIPEKSDYGNLT	LAEDTNIIGSTILT	480
Db	421	TIEVSDKDFKTLCFVQINVIDINDQIPIPEKSDYGNLT	LAEDTNIIGSTILT	480
Qy	481	FTGSSKILYHIIRKDGSEGRGLGVDTPHTNTGYVIKKPLD	FATAAYSNIVFRAENPEPLV	540
Db	481	FTGSSKILYHIIRKDGSEGRGLGVDTPHTNTGYVIKKPLD	FATAAYSNIVFRAENPEPLV	540
Qy	541	FGVKYNASSFAKFTLIVTDVNEAPQFSQHVQAKVSEDVA	IGTKVGNVTAKDPEGLDLSY	600
Db	541	FGVKYNASSFAKFTLIVTDVNEAPQFSQHVQAKVSEDVA	IGTKVGNVTAKDPEGLDLSY	600
Qy	601	SLRGDTRGWLKIDHVTGEIFESVAPLREAGSPYRVQVATE	VGSGSSLSVSSEPHLLIMDV	660
Db	601	SLRGDTRGWLKIDHVTGEIFESVAPLREAGSPYRVQVATE	VGSGSSLSVSSEPHLLIMDV	660
Qy	661	NDNPPRLAKDYGTLFFCHPLSAPGSLIFEATDDQHLFRGP	HFTFSLGSGSLQNDWVEVSK	720
Db	661	NDNPPRLAKDYGTLFFCHPLSAPGSLIFEATDDQHLFRGP	HFTFSLGSGSLQNDWVEVSK	720
Qy	721	INGTHARLSTRHDTDFEERAYVWLIRINDGGRPPLEG	IVSLPTFCSCVSGSCFRPAGHOT	780
Db	721	INGTHARLSTRHDTDFEERAYVWLIRINDGGRPPLEG	IVSLPTFCSCVSGSCFRPAGHOT	780
Qy	781	GIPTVGMAGVILLTTLVIGIILLAVVFIIRIKDKDKDNVES	ASQASVKKPLRS	832
Db	781	GIPTVGMAGVILLTTLVIGIILLAVVFIIRIKDKDKDNVES	ASQASVKKPLRS	832

RESULT 2

RESOLUT 2  
PCT-US02-11475A-1087  
; Sequence 1087, Application PC/TUS0211475A

; GENERAL INFORMATION:

: APPLICANT: Corixa Corporation

APPLICANT: Xu, Jiangchun

APPLICANT: Xu, JIANGCHUN

APPLICANT: LOUES, MICHAEL

; APPLICANT: Benson, David

APPLICANT: Benson, Darin R.  
APPLICANT: Weather, Wadeline

APPLICANT: Meagher, Made.

APPLICANT: Stolk, John A

APPLICANT: Wang, Tongtong

; APPLICANT: Jiang, Yuqiu

; APPLICANT: Smith, Carole

; APPLICANT: King, Gordon I

; APPLICANT: Wang, Aijun

; APPLICANT: Clapper, Jonathan

APPLICANT: Skeiky, Yasir

; APPLICANT: Fanger, Gary I

APPLICANT: Vedvick Thomas

APPLICANT: Carter, Darrick

INVENTOR: CAROL J. BATTEN  
: TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSTICS

TITLE OF INVENTION: COMFOONES

TITLE OF INVENTION: OF COLON

FILE REFERENCE: 210121 47102PC

FILE REFERENCE: 210121.4710ZPC  
: CURRENT APPLICATION NUMBER: PCT/

: CURRENT FILING DATE: 2002  
 : CURRENT AFFILIATION NUMBER:

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1087
; LENGTH: 958
; TYPE: prt
; ORGANISM: Homo sapiens
PCT-US02-11475A-1087

```

Query Match 97.48: Score 4208: DB 1: Length 958:

Query Match	57.4%	Score 4206,
Best Local Similarity	99.48%	Pred No. 0:

BEST LOCAL SIMILARITY 35.4%; PRED. NO: 0;  
Matches 813: Conservative 1: Mismatches 2: Indels 2: Gaps 1:

Qy	17	LATG--YQEGKSGPLKPMTFYSIEGQEPSQIIFOFKANPPAVTVELGTETDNI	74
Db	141	LASGAAQBECKTSGPLKPMTFYSIEGQEPSQIIFOFKANPPAVTVELGTETDNI	74
Qy	75	EGLLYNRALDRSTRSHNLQVAALDANGIIIVGPGPITIEVKDINDNPTFLQSYEGS	134
Db	201	EGLLYNRALDRSTRSHNLQVAALDANGIIIVGPGPITIEVKDINDNPTFLQSYEGS	260
Qy	135	VQNSRPGKPLFVYNATDLDPATPNGLYQYQIVOLPMINNNYMQINNKTAISLTR	194
Db	261	VQNSRPGKPLFVYNATDLDPATPNGLYQYQIVOLPMINNNYMQINNKTAISLTR	320
Qy	195	GSQELNPAKNPSNLYSVKDMGGQSENSESDTTSVDIIIVTENIWKAPXPMVENSTDP	254
Db	321	GSQELNPAKNPSNLYSVKDMGGQSENSESDTTSVDIIIVTENIWKAPXPMVENSTDP	380
Qy	255	HPKITQVRNWDPCAQSLYVDKEKLPFPFSDIOEGDIYVTPQLDREEDKDAYVYAVAKD	314
Db	381	HPKITQVRNWDPCAQSLYVDKEKLPFPFSDIOEGDIYVTPQLDREEDKDAYVYAVAKD	440
Qy	315	EYGPLSYPLEIHHVKVDINDNPPTCPSPTVFEVGENERLGNISIGTLTAHDRDEENTAN	374
Db	441	EYGPLSYPLEIHHVKVDINDNPPTCPSPTVFEVGENERLGNISIGTLTAHDRDEENTAN	500
Qy	375	SFLNRYVEOTPKLPMDGLELIQTYAGMLQAKOSLKKQDTPQYNLTIEVSDKDKFTLCF	434
Db	501	SFLNRYVEOTPKLPMDGLELIQTYAGMLQAKOSLKKQDTPQYNLTIEVSDKDKFTLCF	560
Qy	435	VQINVIDNQIPIFEKSDYGNLTLAEDTNGISTILTIQATDADEPFTGSSKILYHIK	494
Db	561	VQINVIDNQIPIFEKSDYGNLTLAEDTNGISTILTIQATDADEPFTGSSKILYHIK	620
Qy	495	DSEGRGLVDTPHTNTGYVLIKKPLDPETAASNIVFKAENPEPLVFGVKYNASSFAKT	554
Db	621	DSEGRGLVDTPHTNTGYVLIKKPLDPETAASNIVFKAENPEPLVFGVKYNASSFAKT	680
Qy	555	LIVTDVNEAPQFSOHVFOAKVSEDVAIGTKVGNVTAKDPEGLDISYSLRGDTRGWLKIDH	614
Db	681	LIVTDVNEAPQFSOHVFOAKVSEDVAIGTKVGNVTAKDPEGLDISYSLRGDTRGWLKIDH	740
Qy	615	VTGEIFSAPLDREAGSPYRVQVVADEVGSSLSVSEVFHLLIMDVNDNPPRLAKDYTGL	674
Db	741	VTGEIFSAPLDREAGSPYRVQVVADEVGSSLSVSEVFHLLIMDVNDNPPRLAKDYTGL	800
Qy	675	FFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEVSKINGTHARLSRHTD	734
Db	801	FFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEVSKINGTHARLSRHTD	860
Qy	735	FEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCEVSGCFRPAHQGTGPTVGNMAGILLT	794
Db	861	FEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCEVSGCFRPAHQGTGPTVGNMAGILLT	920
Qy	795	TLIVIGIILAVFIRIKKDKGNDESQAQSEVYKPLRS	832
Db	921	TLIVIGIILAVFIRIKKDKGNDESQAQSEVYKPLRS	958

REFSIII.T 3

RESULT 3  
ITS-09-949-016-9669

US-09-949-016-9669 : Sequence 9669 Application IIS/09949016

; sequence 9669, Appl  
; GENERAL INFORMATION:

GENERAL INFORMATION: VENTER, I Craig et al







```

; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079294
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: 60/079556
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/079728
; PRIOR FILING DATE: 1998-03-27
; Remaining Prior Application data re
; NUMBER OF SEQ ID NOS: 246
; SEQ ID NO 188
; LENGTH: 807
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-219-473-188

```

Query Match	21.08;	Score	908.5;	DB	6;	Length	807;
Best Local Similarity	30.24;	Pred.	No. 3.6e-73;				
Matches	237;	Conservative	159;	Mismatches	335;	Indels	53;
Gaps	21;						
QY	62	LTGE	---	TDNI	FI	VERE	-GLLYN
Db							
QY	57	LSGD	SG	KAT	EG	PAM	DP
Db							
QY	117	KDIN	DR	PT	LQ	SK	YEG
Db							
QY	117	KDIN	DR	PT	LQ	SK	YEG
Db							
QY	177	VMY	FQ	IN	KN	K	T
Db							
QY	177	PD	MF	EL	PR	IG	LA
Db							
QY	237	NW	KAP	P	V	E	N
Db							
QY	236	STW	S	L	E	P	I
Db							
QY	297	PLD	R	E	K	D	A
Db							
QY	293	ELD	R	E	A	Q	E
Db							
QY	357	NSI	G	T	T	A	H
Db							
QY	353	TEV	R	S	A	E	D
Db							
QY	415	TPQ	N	L	T	I	E
Db							
QY	473	QAT	D	A	D	-	E
Db							
QY	473	TAD	A	D	L	E	P
Db							
QY	529	I	V	K	A	E	N
Db							
QY	589	TAK	D	P	E	G	L
Db							
QY	587	QPS	D	I	S	R	T
Db							
QY	648	SSV	S	E	F	H	L
Db							
QY	643	---	---	---	---	---	---
Db							
QY	707	LG	-	S	G	S	L
Db							
QY	684	LGP	N	T	V	Q	R
Db							
QY	766	SC	-	V	E	G	S
Db							
QY	741	RCN	V	E	G	O	C
Db							
QY	825	SEV	K	828			

Db 801 VPLK 804

## RESULT 8

US-10-219-477-188  
 ; Sequence 188, Application US/10219477  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Gerritsen, Mary  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Grimaldi, J. Christopher  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Stephan, Jean-Philippe F  
 ; APPLICANT: Watanabe, Colin L.

Query Match	21.08;	Score	908.5;	DB	6;	Length	807;
Best Local Similarity	30.24;	Pred.	No. 3.6e-73;				
Matches	237;	Conservative	159;	Mismatches	335;	Indels	53;
Gaps	21;						
QY	62	LTGE	---	TDNI	FI	VERE	-GLLYN
Db							
QY	57	LSGD	SG	KAT	EG	PAM	DP
Db							
QY	117	KDIN	DR	PT	LQ	SK	YEG
Db							
QY	117	KDIN	DR	PT	LQ	SK	YEG
Db							
QY	177	VMY	FQ	IN	KN	K	T
Db							
QY	177	PD	MF	EL	PR	IG	LA
Db							
QY	237	NW	KAP	P	V	E	N
Db							
QY	236	STW	S	L	E	P	I
Db							
QY	297	PLD	R	E	K	D	A
Db							
QY	293	ELD	R	E	A	Q	E
Db							
QY	357	NSI	G	T	T	A	H
Db							
QY	353	TEV	R	S	A	E	D
Db							
QY	415	TPQ	N	L	T	I	E
Db							
QY	473	QAT	D	A	D	-	E
Db							
QY	473	TAD	A	D	L	E	P
Db							
QY	529	I	V	K	A	E	N
Db							
QY	589	TAK	D	E	P	E	G
Db							
QY	587	QPS	D	I	S	R	T
Db							
QY	648	SSV	S	E	F	H	L
Db							
QY	643	---	---	---	---	---	---
Db							
QY	707	LG	-	S	G	S	L
Db							
QY	766	SC	-	V	E	G	S
Db							
QY	741	RCN	V	E	G	O	C
Db							
QY	825	SEV	K	828			

Query Match	21.0%	Score	908.5	DB	6	Length	807
Best Local Similarity	30.2%	Pred. No.	3.6e-73				
Matches 237	Conservative	159	Mismatches	335			
			Indels	53	Gaps		21

[illegible]



```
Db 293 ELDRQAQAYLLQVRAQNSHGEDYAAPLEHLVLMVMDNDNVIPCPRPDPTVSIPELSPPG 352
QY 357 NSIGTTLTAHRDEENTANSFLNRYVQTPKLPMDG-LFLIQTAYAGMLQAKSLKK-QD 414
Db 353 TEVTRLSAEDADAPGSPNSHVYQLLSPEDGVEGRAFOVDPTSGSVTLGVPLPLRAGON 412
QY 415 TPQYNLTIEVSKD--FKTLCFVOINVIDINDQIPIPEKSDYGNLTAEADTNGSTILTI 472
Db 413 ILLVLAMDLAAGEGFSSCEVEAVTDINDHAPEFITSQIGPISLPEDVEPGTLVAML 472
QY 473 QATDAD-EPFTGSSKIL-YHIKGDSEGLGVDTDPHTNTGYVLIK--KPLDFTAAVSN 528
Db 473 TAIADALEP--AFRLMDFAIERDTEGTGLDWEF--DSGHVRLRCKNLSYEAPASHE 527
QY 529 IVFKAENPEPLVFGVKYNASSFAKTLIVTDVNEAPQFQSHVFOAKVSEDAVIGTKVGNV 588
Db 528 VVVVQSVAKLV-GPGPGCATATVTVLVERVMPKLDQESYEASVPI SAPAGSFLITI 586
QY 589 TAKDPEGLDISYSLRGTRGWLKIDHVTGEIFSVAPLD-REAGSPYRVQVVAIVEVGSGL 647
Db 587 QPSDPTISRTLRFLSVNDSEGLCKIEKFSGEVHTAQSLQGAQPGDITYVLVEAQDTA---- 642
QY 648 SSVSEFHLIIMDVNDNPPRLAKDYTGLEFCHPLSAPGSLIF-EATDDDDQHLFRGPHEFES 706
Db 643 -----LTAPVPSQ-----YLCPTQDQHLVSGPSKDPDLASGHGP-YSFT 683
QY 707 LG-SGSLQNDWEYSKINGTHARLSTRHTDPEERAYVYVLRINDGRRPPEGLISLPVTF 765
Db 684 LGPNPTVQDWRQLQTNGSHAYLTALHWHVEPREHIIPVVVSHNAQ---MWQLLVRVIVC 740
QY 766 SC-VEGSCFRPAGHOTGIPVGMVAGILLTLLVIGIILAVVFIIRIKKDKGNVESAQA 824
Db 741 RCNVEGCMRKVGRMGKMPKLSAVGILVGTVAIGIFILILIEFTHWTMTSRKKDPDQADS 800
QY 825 SEVK 828
Db 801 VPLK 804

RESULT 9
US-10-015-610A-229
; Sequence 229, Application US/10015610A
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2830PIC52
; CURRENT APPLICATION NUMBER: US/10/015, 610A
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/098716
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098723
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098749
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098750
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098803
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098821
```

```
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098843
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/099536
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099596
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099598
; PRIOR FILING DATE: 1998-09-09
; Remaining prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 229
; LENGTH: 807
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-015-610A-229
```

```
Query Match 28.08; Score 908.5; DB 6; Length 807;
Best Local Similarity 30.28; Pred. No. 3.6e-73;
Matches 237; Conservative 159; Mismatches 335; Indels 53; Gaps 21;
```

```
QY 62 LTGE---TDNIFVIERE-GLIYYNRALDRSTRTHNLQVAALDANGLIIVGGVPITIEV 116
Db 57 LSGDSGRATEGPFAMDPSGFLVTRALDRREEQAEYQLQVTLQMDQGHVLMGPQVLVHV 116
QY 117 KDINDNRPTLQSKYEGSVQRNSRPGKPFLLYVNAATDLDPATNGQLYQYIVIQLPINN 176
Db 117 KDENDQVPHESQAIYRARLSRGTRPGIPFLFLEASDRDEPCTANSDLRFHLSQAQPS 176
QY 177 VMFQINNKGTATSLREGSOELNPAKNPSYNLVISVKMGQSGENSEFSDTSTVDIIVTE 236
Db 177 PMFQLEPRLGALALSPKSTSLDHALERTYQLLVQVKMDGQA-SGHQATATVEVSIE 235
QY 237 NIKKAPKPEVMENSTDPHPKIKITQVRWNDPQGAQYSLVDKEKLPREFSIDQEDIVYVQ 296
Db 236 STWVSLEPIHLAENLKVLYPHHMAQVHWSGGDVHYHL---ESHPPGPFVEVNAEGLNYVTR 292
QY 297 PLDREEDKAYVYAVAKDEXGKPLSYPLEIHKVVKDINDNPPTCPSPVTVEFVENERLG 356
Db 293 ELDRQAQAYLLQVRAQNSHGEDYAAPLEHLVLMVMDNDNVIPCPRPDPTVSIPELSPPG 352
QY 357 NSIGTTLTAHRDEENTANSFLNRYVQTPKLPMDG-LFLIQTAYAGMLQAKSLKK-QD 414
Db 353 TEVTRLSAEDADAPGSPNSHVYQLLSPEDGVEGRAFOVDPTSGSVTLGVPLPLRAGON 412
QY 415 TPQYNLTIEVSKD--FKTLCFVOINVIDINDQIPIPEKSDYGNLTAEADTNGSTILTI 472
Db 413 ILLVLAMDLAAGEGFSSCEVEAVTDINDHAPEFITSQIGPISLPEDVEPGTLVAML 472
QY 473 QATDAD-EPFTGSSKIL-YHIKGDSEGLGVDTDPHTNTGYVLIK--KPLDFTAAVSN 528
Db 473 TAIADALEP--AFRLMDFAIERDTEGTGLDWEF--DSGHVRLRCKNLSYEAPASHE 527
QY 529 IVFKAENPEPLVFGVKYNASSFAKTLIVTDVNEAPQFQSHVFOAKVSEDAVIGTKVGNV 588
Db 528 VVVVQSVAKLV-GPGPGCATATVTVLVERVMPKLDQESYEASVPI SAPAGSFLITI 586
QY 589 TAKDPEGLDISYSLRGTRGWLKIDHVTGEIFSVAPLD-REAGSPYRVQVVAIVEVGSGL 647
Db 587 QPSDPTISRTLRFLSVNDSEGLCKIEKFSGEVHTAQSLQGAQPGDITYVLVEAQDTA---- 642
QY 648 SSVSEFHLIIMDVNDNPPRLAKDYTGLEFCHPLSAPGSLIF-EATDDDDQHLFRGPHEFES 706
Db 643 -----LTAPVPSQ-----YLCPTQDQHLVSGPSKDPDLASGHGP-YSFT 683
QY 707 LG-SGSLQNDWEYSKINGTHARLSTRHTDPEERAYVYVLRINDGRRPPEGLISLPVTF 765
Db 684 LGPNPTVQDWRQLQTNGSHAYLTALHWHVEPREHIIPVVVSHNAQ---MWQLLVRVIVC 740
QY 766 SC-VEGSCFRPAGHOTGIPVGMVAGILLTLLVIGIILAVVFIIRIKKDKGNVESAQA 824
Db 741 RCNVEGCMRKVGRMGKMPKLSAVGILVGTVAIGIFILILIEFTHWTMTSRKKDPDQADS 800
```

QY 825 SEVK 828  
:!  
Db 801 VPLK 804

RESULT 10

US-10-226-254A-229  
; Sequence 229, Application US/10226254A  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Botstein, David  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Eaton, Dan I.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Hillan, Kenneth J.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.  
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
; TITLE OF INVENTION: Acids Encoding the Same  
; FILE REFERENCE: P2830PIC68  
; CURRENT APPLICATION NUMBER: US/10/226,254A  
; CURRENT FILING DATE: 2002-08-21  
; PRIOR APPLICATION NUMBER: 60/098716  
; PRIOR FILING DATE: 1998-09-01  
; PRIOR APPLICATION NUMBER: 60/098723  
; PRIOR FILING DATE: 1998-09-01  
; PRIOR APPLICATION NUMBER: 60/098749  
; PRIOR FILING DATE: 1998-09-01  
; PRIOR APPLICATION NUMBER: 60/098750  
; PRIOR FILING DATE: 1998-09-01  
; PRIOR APPLICATION NUMBER: 60/098803  
; PRIOR FILING DATE: 1998-09-02  
; PRIOR APPLICATION NUMBER: 60/098821  
; PRIOR FILING DATE: 1998-09-02  
; PRIOR APPLICATION NUMBER: 60/098843  
; PRIOR FILING DATE: 1998-09-02  
; PRIOR APPLICATION NUMBER: 60/099536  
; PRIOR FILING DATE: 1998-09-09  
; PRIOR APPLICATION NUMBER: 60/099596  
; PRIOR FILING DATE: 1998-09-09  
; PRIOR APPLICATION NUMBER: 60/099598  
; PRIOR FILING DATE: 1998-09-09  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 477  
; SEQ ID NO 229  
; LENGTH: 807  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-226-254A-229

Query Match 21.08; Score 908.5; DB 6; Length 807;  
Best Local Similarity 30.28; Pred. No. 3.6e-73;  
Matches 237; Conservative 159; Mismatches 335; Indels 53; Gaps 21;

QY 62 LTGE---TDNIFVERE-GLLYNRLDRETRSTHNLQVAALDANGIIVEGPPITIEV 116  
:!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!  
Db 57 LSGDSGRATEGFAMDPSGFLVTRALDREGEAYEQVLTLEMDQGHVLMGPPQVLVHV 116  
:!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!  
QY 117 KDINDNRPTFLQSKYGVSRNSRPGKFLVYNATDLPDTPNGQLYQIVQLPMINN 176  
:!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!  
Db 117 KDENDQVHPFSQAIYRRLSRTRPGIFLFLASDRDEPGTANSDLRHLHLSQAQAPS 176  
:!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!  
QY 177 VMYFOINKTKGALSITREGSDELPAKNPNSLVISVKMDGQSENFSDTSDIIVTE 236  
:!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!  
Db 177 PDMFOLEPRLGALALSPRGSTSLDHALERTYQLLVQVVKMDGQA-SGHQATATVEVSTIE 235  
:!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!  
QY 237 NIWAKPKPVENVSTDPHPKIKITQVRNRNDFCAQYSLVDKEKLPREFSIDQEGDIYVTQ 296  
:!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!

Db 236 STWVSLPIHIAENLKVLYPHHMAQVHWSGDDVHYHL---ESHPPGPFVNAEGNLVTR 292  
:!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!  
QY 297 PLDREKDAYVYAVAKDEYKPLSYPLEIHVKYKDIINDNPTCPSPVTVFVEQENERLG 356  
:!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!  
Db 293 ELDRQAQAEYLLQVRAQNSHGEDYAAPLEHLVLMVDENDNVPICPPRDPTVSIPELSPG 352  
:!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!  
QY 357 NSIGTTLAHRDRENTANSFLNRIVEQTPKLPMDG-LFLIQTIVAGMLQAKOSLKK-QD 414  
:!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!  
Db 353 TEVTRLSAEDADAPGSPNSHVYQLLSPEDGEGVRAFOVDPTSGSVTLGLVPLRAGON 412  
:!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!  
QY 415 TPOYNLTIEVSKD--PKTLCFQVINVIDNQIPIEKSQDYGNLTAEOTNIGSTILTI 472  
:!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!  
Db 413 ILLVLANDLAGAEGGFSSTCEVEVATDINDHAPEITSOIGPISLPEDEPGTLVAML 472  
:!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!  
QY 473 QATDAD-EPFTGSSKIL-YHIKGDSEGRGLVDTPHTNTGYVIK--KPLDFTAAVSN 528  
:!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!  
Db 473 TAIDADLEP--AFRLMDFAIERGDTEGTGLDWEF--DSGHVRLRLCKNLSEAAAPSHE 527  
:!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!  
QY 529 IVFKAENPEPLVFGVKYNASSFAKFTLLIVTDVNEAPQFSQHVFOAKYSEDVAICTKGVN 588  
:!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!  
Db 528 VVVVQSAVLV-GPGPGGATATVTLVERVMPKLDQESYEASVPISAPAGSEFLTI 586  
:!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!  
QY 589 TAKDPEGLDISYSLRGDTRGWLKIDHVTGEIFSVAPLD-REAGSPYRVQVVATEVGGSSL 647  
:!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!  
Db 587 QSPDISRTLRFLSVNDSEGLCKIEKFSGEVHTAQSLOGAQPQDITYVLVEAQDTA---- 642  
:!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!  
QY 648 SSVSEFHLILMDVNDNPPRLAKOYTGTLFFCHPLSAPGSLIF-EATDDDHHLFRGPHFTFS 706  
:!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!  
Db 643 -----LTLAPVPSQ-----YLCPTQDQHLIVSGPSKDPDLASCHGP-YSFT 683  
:!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!  
QY 707 LG-SGSLONDWEVSKINGTHARLSTRHTDDEERAYVVLIRINDGRPPLEGIVSLPVTFC 765  
:!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!  
Db 684 LGPNPTVQRDWRMLQTLNGSHAYLTLALHWEPREHILIPVVVSHNAQ----MKQLLVRLVLC 740  
:!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!  
QY 766 SC-VEGSCFRPAGHTGIPVTGMAVGILLTLLVIGIILAVVFIIRIKDKDKDNVESQA 824  
:!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!  
Db 741 RCNVEGQCMRKVGRMGKMPKLSAVGILVGLVAIGLFLIFLTHWTMSRKKKDPQPADS 800  
:!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!: :!  
QY 825 SEVK 828  
:!  
Db 801 VPLK 804

RESULT 11

US-10-017-253A-229  
; Sequence 229, Application US/10017253A  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Botstein, David  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Eaton, Dan I.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Hillan, Kenneth J.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.  
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
; TITLE OF INVENTION: Acids Encoding the Same  
; FILE REFERENCE: P2830PIC62  
; CURRENT APPLICATION NUMBER: US/10/017,253A  
; CURRENT FILING DATE: 2001-12-13  
; PRIOR APPLICATION NUMBER: 60/098716  
; PRIOR FILING DATE: 1998-09-01  
; PRIOR APPLICATION NUMBER: 60/098723  
; PRIOR FILING DATE: 1998-09-01  
; PRIOR APPLICATION NUMBER: 60/098749  
; PRIOR FILING DATE: 1998-09-01

```
; PRIOR APPLICATION NUMBER: 60/098750
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098803
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098821
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098843
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/099536
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099596
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099598
; PRIOR FILING DATE: 1998-09-09
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 229
; LENGTH: 807
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-017-253A-229

Query Match          21.0%; Score 908.5; DB 6; Length 807;
Best Local Similarity 30.2%; Pred. No. 3.6e-73;
Matches 237; Conservative 159; Mismatches 335; Indels 53; Gaps 21;

Qy 62 LTGE-----TDNIFVIERE-GLLYNRLDRETRSHNLQVAALDANGIIVEGVPITIEV 116
Db 57 LSGDGGKATEGPFAMDPSGFLVTRALDREAEQYQVLTLEMQDGHVWGQPVLVHV 116
Qy 117 KDINDNRFTTQSKYEGSVQRNSRPGKPFLLVYNATLDDPATPNQOLYQIVQLPMNN 176
Db 117 KDENDQVPHFSQAIYRARLSRGTRPGIPFLFLEASDRDEPGTANSDLRPHILSQAPQPS 176
Qy 177 VMYFOINKNTGAISLTRGSOELNPAKNPSYNLVSVKMDMGOSSEFSDTTSVDIIVTE 236
Db 177 PMFOLEPRLGALSPKGSSTLDHALERTYQLLVQVQKMDGQA-SGHOATATVEVSIE 235
Qy 237 NIWAKPKVEMVNSTDPHTKITQVRWNPQGAQYSLVDKELPRFPESIDQEGDIYVYQ 296
Db 236 STWVSLEPIHLAENLVLYPHMAQVHWSGGDVHYHL---ESHPPGPFVNAEGNLYYTR 292
Qy 297 PLDREKDAYVYFAVAKDEYKPLSYPLEIHVKVKDINDNPPTCSPVTVFEVQNERLG 356
Db 293 ELDRQAQAEYLLQVRAQNSHGEDIAAPLEHLVLMENDNVNPPICPRDPTVSIPELSPG 352
Qy 357 NSICTLTAHDREENTANSFLNYRIVEQTPKLPMDG-LFLIQTAYAGMLQAKQSLK-OD 414
Db 353 TEVTRLSAEDADAPGSPNSHVYQLLSPEDGVEGRAFDVDPDTSVTLGLVPLRAGON 412
Qy 415 TPQYNLTIEVSDKO--FKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILT 472
Db 413 ILLVLMDLAGAEGGFSTCEVEVAVVDINDHAFETISQIGPISLPEDEVPGLVAML 472
Qy 473 QATDAD-EPFTGSSKIL-YHIKGDSEGRGLGVDTPHTNTGYVIK--KPLDFETAAYSN 528
Db 473 TAIDADLEP---AFRLMDFAIERGDTGTGFLDWEF--DSGHVRLRLCKNLSYEAPSHS 527
Qy 529 IVFRAENPEPLVFGVKYNASSFAKETLLIVDVNEAPQSFQHVQAKVSEDVAIGTKVGNV 588
Db 528 VVVVVQSVAKVL-GPGPGGATATVTLVVERVMPKPKLDQESYEASVPISAPAGSFLITI 586
Qy 589 TAKDPEGDIDSVLRGDFRWGLKIDHVTGEIFSVAPLO-REAGSPYRQVQVATEVGGSSL 647
Db 587 QPSDPISTLRFSLVNDSEGNLCIEKFSGEVHTAQSLQAGQPGDTYTVLVEAQDTA---- 642
Qy 648 SSVSEFHILMDVNDNPPRLAKDYTGFLFCHPLSAPGSLIF-EAFTDDQHLFRGPHFTFS 706
Db 643 -----LTLAPVPSQ-----YLCPTRODHGLIVSGPSKDPDLASGHGP-YSET 683
Qy 707 LG-SSGLONDNEVSKINGTHARLSTRHRTDFERAYVWLIRINDGGRPLEGIVSLPVTF 765
Db 684 LGPNTVQDRMLQTLNGSHAYLTALHWEVPEPREHIIPVVVSHNAQ---MMQLLVRVIVC 740
```

```
Qy 766 SC-VEGSCFRPAGHOTGPTVGMVGIILTTLLVIGIILVAVFIRIKKDKGRDNVESQA 824
Db 741 RCNVEGQCMRKVGRMKMPTKLSAVGILVGLVLAIGLIFLIFHTHTWSRKKDPDPQADS 800
Qy 825 SEVK 828
Db 801 VPLK 804

RESULT 12
US-10-063-557-98
; Sequence 98, Application US/10063557
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Eaton, Dan L.
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: GNE.3230R1C39
; CURRENT APPLICATION NUMBER: US/10/063.557
; CURRENT FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: PCT/US00/23328
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: PCT/US99/20111
; PRIOR FILING DATE: 1999-09-01
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: US 60/169,495
; PRIOR FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: US 60/170,262
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: US 60/175,481
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: PCT/US00/04341
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: PCT/US00/04342
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/05601
; PRIOR FILING DATE: 2000-03-01
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 98
; LENGTH: 807
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-063-557-98

Query Match          21.0%; Score 908.5; DB 6; Length 807;
Best Local Similarity 30.2%; Pred. No. 3.6e-73;
Matches 237; Conservative 159; Mismatches 335; Indels 53; Gaps 21;

Qy 62 LTGE-----TDNIFVIERE-GLLYNRLDRETRSHNLQVAALDANGIIVEGVPITIEV 116
Db 57 LSGDGGKATEGPFAMDPSGFLVTRALDREAEQYQVLTLEMQDGHVWGQPVLVHV 116
Qy 117 KDINDNRFTTQSKYEGSVQRNSRPGKPFLLVYNATLDDPATPNQOLYQIVQLPMNN 176
Db 117 KDENDQVPHFSQAIYRARLSRGTRPGIPFLFLEASDRDEPGTANSDLRPHILSQAPQPS 176
Qy 177 VMYFOINKNTGAISLTRGSOELNPAKNPSYNLVSVKMDMGOSSEFSDTTSVDIIVTE 236
Db 177 PMFOLEPRLGALSPKGSSTLDHALERTYQLLVQVQKMDGQA-SGHOATATVEVSIE 235
```

```
Qy 237 NIWAKPVPVEMVENSTDPHPKIKITQVRWNPDPGAQYSLVDKEKLPFRPFSDIDQEGDIYVTO 296
Db 236 STWVLEPIHLAENLKVLYPHHMAQVHWSGVDVHYHL---ESHPPGPFVEVNAEGNLVYTR 292
Qy 297 PLDREKDAYVFAVAKDEYKPLSYPLEIHKVVKDINDNPPCTCPSVTVFEQVNERLG 356
Db 293 ELDRQAQVLLQVRAQNSHGEDYAAPLEHLVLMVDENDNVPICPPRDPVTSIPELSPPG 352
Qy 357 NSIGTLTAHRDRENTANSFLNRIVEQTPKPLMDG-LFLIQTAYAGMLQAKQSLKK-QD 414
Db 353 TEVTRLSAEDADAPGSPNSHVYQLLSPEPDGVEGRAFOVDPTSGSVTLGLVPLRAGQN 412
Qy 415 TPQYNLTIEVSDKD--FKTLCFVOINVIDINDQIPFEKSDYGNLTAEADNIGSTILTI 472
Db 413 ILLVLVLAHDLAGAGGFSSTCEVEAVTDINDHAPEFITSQIGPISLPEDVEPGLVAML 472
Qy 473 QATDAD-EPFTGSSKIL-YHIKGDSEGRGLVDTPHTNTGYVLIK--KPLDFETAASN 588
Db 528 VVVVQSVAKLV-GPGPGGATATVTLVERVMPKLDQESYEASVPISAPAGSFLITI 586
Qy 589 TAKDEGLDISYLRGDTGRHLKIDHVTGEIFSVAPLD-REAGSPYRVQVATEVGGSSL 647
Db 587 QPSDPISRTLRFSLVNDSEGWLCEKFSGEVHTAQSLQGAQPGDTYTVLYEAQDTA---- 642
Qy 648 SSVSEFHLILMDVNDNPPRLAKDVTGLFFCHPLSAPGSLIF-EATDDDDQHLFRGPHFTFS 706
Db 643 -----LTLAPVPSQ-----YLCPTRODHGLIVSGPKDPLASGHGP-VSFT 683
Qy 707 LG-SGSLQNDWEVSKINGTHARLSTRHTDFEERAYVYVLLIRINDGRRPPLGIVSLPVTFC 765
Db 684 LGPNPTVQRDRLQTLNGSHAYTLALHWVPREHIIIPVVVSHNAQ---MWQLLVRVIVC 740
Qy 766 SC-VEGSCFRPAGHOTGTPVCMVAGILLTLLVIGIILAVVFIKDKCKDNVESQAQ 824
Db 741 RCNVEGQCMRKVGRMKGMPTKLSAVGILVGLTVAIGIFLILIFTHWTMSRKDKDPPDADS 800
Qy 825 SEVK 828
Db 801 VPLK 804
```

```
RESULT 13
US-09-724-676-54141
; Sequence 54141, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724.676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 54141
; LENGTH: 664
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-54141

Query Match 17.2%; Score 743.5; DB 5; Length 664;
Best Local Similarity 31.4%; Pred. No. 2.4e-58;
Matches 181; Conservative 125; Mismatches 248; Indels 23; Gaps 13;
```

```
Qy 62 LTGE-----TDNIFVIERE-GLLIYNNRDLRETRSTHNNQVAALDANGIIVEGVPVITIEV 116
Db 57 LSGSGKATEGPFAMDPDSGFLLVTRALDREQAQYQLQVTLQEMODGHVLMGPPQVPLVHV 116
Qy 117 KDINDNRRPTFQSKYEGSVRQNSRPGKPLFYNNATDLDPPATPNQOLYYQIVQLPMINN 176
Db 117 KDENDOVPHFSQAIIYRARLSRGTREGIPFLFEASDRDEPGTANSDLRHFHILSQAPAPS 176
```

```
Qy 177 VMYFOINNKTAISLTRREGSQELNPAKNPSYNLVIVSKMDGQSGSENSFSDTTSVDIIIVTE 236
Db 177 PDMFQLEPRLGALALSPKGSTSLDHALERTYQLLVQVKMDGQA-SGHOATATVEVSIIIE 235
Qy 237 NIWAKPVPVEMVENSTDPHPKIKITQVRWNPDPGAQYSLVDKEKLPFRPFSDIDQEGDIYVTO 296
Db 236 STWVLEPIHLAENLKVLYPHHMAQVHWSGVDVHYHL---ESHPPGPFVEVNAEGNLVYTR 292
Qy 297 PLDREKDAYVFAVAKDEYKPLSYPLEIHKVVKDINDNPPCTCPSVTVFEQVNERLG 356
Db 293 ELDRQAQVLLQVRAQNSHGEDYAAPLEHLVLMVDENDNVPICPPRDPVTSIPELSPPG 352
Qy 357 NSIGTLTAHRDRENTANSFLNRIVEQTPKPLMDG-LFLIQTAYAGMLQAKQSLKK-QD 414
Db 353 TEVTRLSAEDADAPGSPNSHVYQLLSPEPDGVEGRAFOVDPTSGSVTLGLVPLRAGQN 412
Qy 415 TPQYNLTIEVSDKD--FKTLCFVOINVIDINDQIPFEKSDYGNLTAEADNIGSTILTI 472
Db 413 ILLVLVLAHDLAGAGGFSSTCEVEAVTDINDHAPEFITSQIGPISLPEDVEPGLVAML 472
Qy 473 QATDAD-EPFTGSSKIL-YHIKGDSEGRGLVDTPHTNTGYVLIK--KPLDFETAASN 528
Db 473 TAIDADLEP---APRLMDFAIERGDTGTEGFLDWEF--DSGHVRLRCKNLSEAAPSHE 527
Qy 529 IVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPOSQHVFOAKYSEDVAIGTKVGNV 588
Db 528 VVVVQSVAKLV-GPGPGGATATVTLVERVMPKLDQESYEASVPISAPAGSFLITI 586
Qy 589 TAKDEGLDISYLRGDTGRHLKIDHVTGEIFSVAPL 625
Db 587 QPSDPISRTLRFSLVNDSEGWLCEKFSGEVHTAQSL 623
```

```
Query Match 17.2%; Score 743.5; DB 5; Length 664;
Best Local Similarity 31.4%; Pred. No. 2.4e-58;
Matches 181; Conservative 125; Mismatches 248; Indels 23; Gaps 13;
```

```
Qy 62 LTGE-----TDNIFVIERE-GLLIYNNRDLRETRSTHNNQVAALDANGIIVEGVPVITIEV 116
Db 57 LSGSGKATEGPFAMDPDSGFLLVTRALDREQAQYQLQVTLQEMODGHVLMGPPQVPLVHV 116
Qy 117 KDINDNRRPTFQSKYEGSVRQNSRPGKPLFYNNATDLDPPATPNQOLYYQIVQLPMINN 176
Db 117 KDENDOVPHFSQAIIYRARLSRGTREGIPFLFEASDRDEPGTANSDLRHFHILSQAPAPS 176
Qy 177 VMYFOINNKTAISLTRREGSQELNPAKNPSYNLVIVSKMDGQSGSENSFSDTTSVDIIIVTE 236
Db 177 PDMFQLEPRLGALALSPKGSTSLDHALERTYQLLVQVKMDGQA-SGHOATATVEVSIIIE 235
Qy 237 NIWAKPVPVEMVENSTDPHPKIKITQVRWNPDPGAQYSLVDKEKLPFRPFSDIDQEGDIYVTO 296
Db 236 STWVLEPIHLAENLKVLYPHHMAQVHWSGVDVHYHL---ESHPPGPFVEVNAEGNLVYTR 292
Qy 297 PLDREKDAYVFAVAKDEYKPLSYPLEIHKVVKDINDNPPCTCPSVTVFEQVNERLG 356
Db 293 ELDRQAQVLLQVRAQNSHGEDYAAPLEHLVLMVDENDNVPICPPRDPVTSIPELSPPG 352
```

```
Db 293 ELBREAQAYLLQVRAQNSHGEDIAAPLEHLVLMVDNDNVPCPPRDPRTVSIPELSPPG 352
QY 357 NSICTLTAHRDENTANSEFLNRYVQTPKLPMDG--LFLIQTAYAGMLQAKOSLKK--QD 414
Db 353 TEVTRLSAEDADAPGSPNSHVQQLSPEDGVEGRAFOVDTPSGSVTLGVPLPRAGQN 412
QY 415 TPQYNLFIEVSKD--FKTLCFVQINVIDINDOIQIPEKSDYGNLTLAEDTNIGSTILTI 472
Db 413 ILLVLAMDLAAGEGFSSTCEVEVAVTINDHAPEFITSOIGPISLPEDVEPGTLVAML 472
QY 473 QATDAD--EPTGSSKIL--YHIIGKDSGRGLGVDTDPHTNTGYVIK--KPLDRETAASN 528
Db 473 TADADLEP---APRLMFAIERGDTGTEGLDWEF--DSGHVRLRCKNLSYEAAPSHE 527
QY 529 IVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPOFSQHVFOAKVSEDVAIGTKVGNV 588
Db 528 VVVVQSVAKLV--GPGPGCATATVTVLVERVMPKLDQESYEASVPIAPAGSFLITI 586
QY 589 TAKDPEGLDISYLRGTRGWLKIDHVTGRTFSVAPL 625
Db 587 QPSDPISTLRFSLVNDSEGLCKIEKFSGEVHTAQL 623

RESULT 15
US-60-452-680-19984
; Sequence 19984, Application US/60452680
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: GRUPE, Andrew
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001450
; CURRENT APPLICATION NUMBER: US/60/452, 680
; CURRENT FILING DATE: 2003-03-07
; NUMBER OF SEQ ID NOS: 116213
; SOFTWARE: fastseq for Windows Version 4.0
; SEQ ID NO 19984
; LENGTH: 786
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-452-680-19984

Query Match 16.2%; Score 699.5; DB 7; Length 786;
Best Local Similarity 27.4%; Pred. No. 3.3e-54;
Matches 210; Conservative 122; Mismatches 331; Indels 103; Gaps 21;

QY 117 KDINDNPTFLQSKYEGSVRQNSRPGKPFYLVNATDLDPATNGQLYQYIQLPMINN 176
Db 47 KDHEGQP--LLNVKF-----SNCNGKRKVQYESSEPADFKVDEGMYV--AVRSFPLSE 98
QY 177 VMYFQINNKTCATSLTREGSQELNPAKNPSYNLVISVKMGQSENFSFDTTSVDIIV-- 234
Db 99 HAKELIY-----AQDKETQEKQVAVKLSLKT--LTEESVKESAEEIVEFP 144
QY 235 -----TENIWKAPKPVEMENSTDPHPKITQVRWN-----DPGAQ 270
Db 145 RQESKSHGLQROKRDWVIP--PINLPENSRGPPFOELVRIKSDRKNLSLRYSVTGPCA- 202
QY 271 YSLVDKEKLPFPFSDIQ--EGDIYVTOPLDREKDAYVFAVAKDEYCKPLSPLEIHKV 329
Db 203 -----DQPTGTPIIIPISGQSVTKPLDREQIAREHLRAHAYVDINGNQVENPIDIVN 256
QY 330 VKOINDNPPTCPSVTVFEVQENRGLNSIGTLTAHREDENTANSFLNRYRIVEOTPKLP 389
Db 257 VIDMNDNRPEELHGVWNGTVPEGSKPQTYVMTVTAIDADDPNALNGLMYRIRVISOAFSTP 316
QY 390 MDGLFLIOTYAGMLQAKOSLKKQDTPQYNLTIEVSKD-----FKTLCFVQINVIDIN 443
Db 317 SPNFTINNETGDIITVAAGLDREKVQOYTLIIQATDMEGNPTYGLSNTATAVITVTDVN 376
QY 444 DOIPIP-EKSDYGNLTLAEDTNIGSTILTIQATDADPEFTGSSKILYHIKGDSEGRIGV 502
Db 377 DNPPEFTAMTYGSEV---PENRVDIIIVANLTVTDKQDPHTPAMNAVYRISGGDPTGRFAI 433
```

```
QY 503 DTDPHNTGTGYIIKKPLDDETAASNIIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNE 562
Db 434 QTDPNSENDGLTVVVKPIDFETNRMFVLTVAAENQVPLAKIQHPPOSTATVSVTIDVNE 493
QY 563 APOFSQHVFOAKVSEDVAIGTKVGNVTAKDPEGL---DISYSLRGDTRGWLKIDHVTGEI 619
Db 494 NPYFAPNPKTIROEBGLHAGTMTLTTTAODPDYMOQNIRYTKLSDPANMLKIDPVANGOI 553
QY 620 FSVAPLDREA---GSPYRVQVVAATEVGGSSLSVSEFHLILMDVNDNPPRLAKDYTGLF 675
Db 554 TTIATVDRESPNVKNNIYNATFLASDNGIIPMSGCTGLQIYLLDINDNAPQVL----- 606
QY 676 FCHPLSA-----PGSLIFEATDDDDQHLFRGPH--FTFSLGSGSLQNDWEVSKINGTHAR 727
Db 607 ---POEAETCETPDPSINITALDYDIDNAGPFAFDPLPSPTIKRNVITIRLNGDFAQ 663
QY 728 LSTRHTDFEERAYVLLIRINDGGRPPLEGIVSLPVTFCSC--VEGSCFRPAGHQGTPTV- 785
Db 664 LNLKIKFLEAGIYEVPIIITDSGNPPKSNISILRVKVCQDCSDGDC-----TDVDRIV 716
QY 786 --GMAYGILLTLLVIGIILAVVFIRIKKDKGKDNVESQAQASEVKP 829
Db 717 GAGLGTGAIITAILLCIIILILJLMLFVVMKRRDKROAKOLLIDP 762
```

Search completed: April 2, 2003, 16:04:47  
Job time : 95.1781 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4\_p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 2, 2003, 15:50:19 ; Search time 30.3927 Seconds  
(without alignments)  
2631.680 Million cell updates/sec

Title: US-09-079-678-178  
Perfect score: 4321  
Sequence: 1 MILQAHLSLCLMLYLATG.....DKGKNVESQAQSEVKPLRS 832

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4303	99.6	832	2 S55396	Li-cadherin - huma
2	3414	79.0	827	2 A53954	Li-cadherin precu
3	954	22.1	829	2 I46536	Ksp-cadherin - rab
4	715	16.5	913	1 IJCHCR	R-cadherin precurs
5	705	16.3	913	1 A47543	R-cadherin precurs
6	699.5	16.2	906	1 IJHUCN	cadherin 2 precurs
7	696	16.1	912	1 IJHUCN	N-cadherin precurs
8	694.5	16.1	906	1 IJMSEN	N-cadherin precurs
9	688.5	15.9	877	1 IJBOCN	N-cadherin precurs
10	684.5	15.8	916	2 C38992	cadherin 4 precurs
11	677	15.7	783	2 I50116	N-cadherin precurs
12	671.5	15.5	826	2 B55363	desmocollin, type
13	671.5	15.5	896	2 A55363	desmocollin, type
14	667.5	15.4	847	1 IJHUDA	desmocollin 3b pre
15	667.5	15.4	901	1 IJHUDA	desmocollin 3a pre
16	663.5	15.4	906	1 IJXLC2	N-cadherin 2 precu
17	661	15.3	896	2 I45858	desmocollin - bovi
18	657	15.2	713	2 B38992	cadherin 13 precu
19	653.5	15.1	809	1 IJBOOD	desmocollin 2b pre
20	653.5	15.1	863	1 IJBOOC	desmocollin 2a pre
21	647.5	15.0	905	1 IJXLC1	N-cadherin 1 precu
22	634	14.7	887	1 IJCHCL	E-cadherin precurs
23	630.5	14.6	712	1 IJMST	T-cadherin precurs
24	630.5	14.6	717	2 I51206	T-cadherin 2 - chi
25	609	14.1	884	2 S34438	uvomorulin - mouse
26	608	14.1	790	2 G02678	cadherin-14 - huma
27	604	14.0	884	1 IJMSE	E-cadherin precurs
28	597.5	13.8	732	1 IJCHCB	B-cadherin precurs
29	591	13.7	882	1 IJHUCE	cadherin 1 precurs

ALIGNMENTS

RESULT 1

S55396

Li-cadherin - human

C:Species: Homo sapiens (man)

C>Date: 15-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 31-Mar-2000

C:Accession: S55396

R:Boettinger, A.; Kreft, B.; Fieger, C.; Dlouhy, B.; Berndorff, D.; Goessner, R.; Tau

submitted to the EMBL Data Library, December 1994

A:Description: Molecular cloning of human Li-cadherin:evidence for a novel type of ca

A:Reference number: S55396

A:Accession: S55396

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-832 <BOE>

A:Cross-references: EMBL:X83228; MID:g854174; PIDN:CAA58231.1; PID:g854175

C:Superfamily: cadherin; cadherin repeat homology

F:455-566/Domain: cadherin repeat homology <CR3>

Query Match 99.6% Score 4303; DB 2; Length 832;

Best Local Similarity 99.5%; Pred. No. 1.3e+262;

Matches 828; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy	1	MILQAHLSLCLMLYLATGYGQEGKFSGLPKPMTFSIYEGQEPSQIIQFKANPPAVTF	60
Db	1	MILQAHLSLCLMLYLATGYGQEGKFSGLPKPMTFSIYEGQEPSQIIQFKANPPAVTF	60
Qy	61	ELTGETDNIFVIERGLLYNRLDRETRSTHNLQVAALDANGIIVEGVPITIEVKDIN	120
Db	61	ELTGETDNIFVIERGLLYNRLDRETRSTHNLQVAALDANGIIVEGVPITIEVKDIN	120
Qy	121	DNRPTELOSKYEGSVRONSRPKPFLYVYNATDLDPPATNGOLYYQIQLPMINNVMVF	180
Db	121	DNRPTELOSKYEGSVRONSRPKPFLYVYNATDLDPPATNGOLYYQIQLPMINNVMVF	180
Qy	181	QINNKTAISLTREGSQELNPAKNPSYNLVLSVKDMGQSENFSDDTTSVDIIVTENIWK	240
Db	181	QINNKTAISLTREGSQELNPAKNPSYNLVLSVKDMGQSENFSDDTTSVDIIVTENIWK	240
Qy	241	APKPVEMVENSTDPHPKITQVRWNDPGAQYSLVDKEKLPREPFSIDEGDIYVTPQLDR	300
Db	241	APKPVEMVENSTDPHPKITQVRWNDPGAQYSLVDKEKLPREPFSIDEGDIYVTPQLDR	300
Qy	301	EEDDAYVFYAVAKDEYKGPLSPLEIHVKVDINDNPPTCPSPVTVFVQENRIGNSIG	360
Db	301	EEDDAYVFYAVAKDEYKGPLSPLEIHVKVDINDNPPTCPSPVTVFVQENRIGNSIG	360
Qy	361	TLTAHDEEENTANSFLNRYIVEQTPLKPMDBGLFIQTYAGMLQAKSLKKQDTPQYNL	420
Db	361	TLTAHDEEENTANSFLNRYIVEQTPLKPMDBGLFIQTYAGMLQAKSLKKQDTPQYNL	420
Qy	421	TIEVSKDKFVKLCFQVINVDINDOIPIFEKSDYGNLFLAEDTNGISTILTIQATDAEP	480
Db	421	TIEVSKDKFVKLCFQVINVDINDOIPIFEKSDYGNLFLAEDTNGISTILTIQATDAEP	480

Db 421 TIEVSDKDFKTLFCVQINVIDNDQTPIFEKSDYGNLTAEEDTNGSTILTITQATDADEP 480  
Qy 481 FTGSSKILYHIIGDSEGRGLGVDTPHNTGYVVIKKPLDPETAAVSNIVKAEENPEPLV 540  
Db 481 FTGSSKILYHIIGDSEGRGLGVDTPHNTGYVVIKKPLDPETAAVSNIVKAEENPEPLV 540  
Qy 541 FGVKNASSFAKFTLIVTDVNEAPQFSOHVFOAKVSEDAVGTGVNVTAKDPEGLDISY 600  
Db 541 FGVKNASSFAKFTLIVTDVNEAPQFSOHVFOAKVSEDAVGTGVNVTAKDPEGLDISY 600  
Qy 601 SLRGDTRGLWKIDHVTGEIFSVAPLDREAGSPYRVQVVADEVGSSLSVSEFHLILMDV 660  
Db 601 SLRGDTRGLWKIDHVTGEIFSVAPLDREAGSPYRVQVVADEVGSSLSVSEFHLILMDV 660  
Qy 661 NDNPRLAKDYGTLFFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEYSK 720  
Db 661 NDNPRLAKDYGTLFFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEYSK 720  
Qy 721 INGHARLSTRHTDFEERAYVVLIRINDGRRPPLGIVSLVPTFCSCVEGSCFRPAGHOT 780  
Db 721 INGHARLSTRHTDFEERAYVVLIRINDGRRPPLGIVSLVPTFCSCVEGSCFRPAGHOT 780  
Qy 781 GIPTVGMAGVILLTLVIGIILAVVFIIRIKKDKGNVESQAQASEVKPLRS 832  
Db 781 GIPTVGMAGVILLTLVIGIILAVVFIIRIKKDKGNVESQAQASEVKPLRS 832

RESULT 2  
A53954  
LI-cadherin precursor - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 07-Oct-1994 #sequence\_revision 07-Oct-1994 #text\_change 31-Mar-2000  
R:Berndorff, D.; Gessner, R.; Kreft, B.; Schnoy, N.; Lajous-Petier, A.M.; Loch, N.; Reut  
J. Cell Biol. 125, 1353-1369, 1994  
A:Title: Liver-intestine cadherin: molecular cloning and characterization of a novel Ca  
A:Reference number: A53954; MUID:94266966; PMID:8207063  
A:Accession: A53954  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-827 <BR>  
A:Cross-references: GB:X78997; NID:g505562; PIDN:CAA55631.1; PID:g505563  
C:Superfamily: cadherin; cadherin repeat homology  
C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; intestine; liver;  
F:1-28/Domain: signal sequence #status predicted <SIG>  
F:29-827/Product: LI-cadherin #status predicted <MAT>  
F:454-565/Domain: cadherin repeat homology <CR3>

Query Match 79.0%; Score 3414; DB 2; Length 827;  
Best Local Similarity 78.8%; Pred. No. 1e-206;  
Matches 655; Conservative 70; Mismatches 102; Indels 4; Gaps 1;

Qy 2 ILQAHLSICLLMLYLANGYGEKGFSGPLKPMFTSIVEGEPISOIIFQFRANPAVTFE 61  
Db 1 MYSQAHLFCLLTLLTYLGAYGEGFSGPLKPMFTSIVEGEPISOIIFQFRANPAVTFE 60

Qy 62 LTGETDNFIVIEREGLLYNRALDRETRSTHNLQVAALDANGIIVEGVPVITIEVKDIND 121  
Db 61 LTGETDGIFKIEKGLLYHTVLDRETRAVHHLQALDQSGAIVDGPVITIEVKDIND 120

Qy 122 NRPTFQSKYEGSVQRNRPCKPFLYVNVATDLDPATPNQLYQIIVTOLPMNNVYFQ 181  
Db 121 NRPTFQKYEKSVQRNRPCKPFLYVNVATDLDPATPNQLYQIIVTOLPMNNVYFQ 180

Qy 182 LNNKTAISLREGSOELNPAKNPSYNLVSVKDMGGOSENFSFDTSDVITENIWA 241  
Db 181 IDNKTAISLREGSOELNPAKNPSYNLVSVKDMGGOSENFSFDTSDVITENIWA 240

Qy 242 PKPVMENSTDPHPKIKITQVRWNPDAQYSLVDKEKLPFRPFSDQEGDIYVTPDLRE 301  
Db 241 PEPVEIRENLDPHPKIKITQVRWNPDAQYSLVDKEKLPFRPFSDQEGDIYVTPDLRE 300

Qy 302 EKDAYVFVAVAKDEYKGPLSVPLFIHVKKVDINDNPPTCPSPVTVFVEQNERLGNISGT 361  
Db 301 EKDSHVFATAKDENGAPLAYPLEIRVKVIDINDNPPTCLSQVTVFVEQNEVLGSSIGI 360  
Qy 362 LTAHREDEENTANSFLNRYRIVEQTPKLPMDGLFLIOTYAGMLQAKOSLKKQDTPQYNLT 421  
Db 361 FAHDMDEANNINSILKYLVDQTPKVPSEDFLIDEYGGKVLGKRSLLKKQDQSPQYNLT 420  
Qy 422 IEVSKDFKTLCTFQVQINVIDNDQIPIPEKSDYGNLTAEEDTNGSTILTITQATDADEPF 481  
Db 421 VEVSIDDFKTLCSLQVNVINDNDQIPIPEKSDYGNLTAEEDTNGSTILTITQATDADEPF 480  
Qy 482 TGSCKILYHIIGDSEGRGLGVDTPHNTGYVVIKKPLDPETAAVSNIVKAEENPEPLV 541  
Db 481 TGSCKILYHIIGDSEGRGLGVDTPHNTGYVVIKKPLDPETAAVSNIVKAEENPEPLV 540  
Qy 542 GVKYNASSFAKFTLIVTDVNEAPQFSOHVFOAKVSEDAVGTGVNVTAKDPEGLDISY 601  
Db 541 GIEYNASSFAKFTLIVTDVNEAPQFSOHVFOAKVSEDAVGTGVNVTAKDPEGLDISY 600  
Qy 602 LRGDTRGLWKIDHVTGEIFSVAPLDREAGSPYRVQVVADEVGSSLSVSEFHLILMDV 661  
Db 601 LKDNKRWGLKIDSVTGDIFFSTAPLDRETESYRVQVVADEVGSSLSVSEFHLILMDV 660  
Qy 662 DNPPRLAKDYGTLFFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEYSK 721  
Db 661 DNPPRLAKDYGTLFFCHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEYSK 720  
Qy 722 NGHARLSTRHTDFEERAYVVLIRINDGRRPPLGIVSLVPTFCSCVEGSCFRPAGHOT 781  
Db 721 NGHARLSTRHTDFEERAYVVLIRINDGRRPPLGIVSLVPTFCSCVEGSCFRPAGHOT 780  
Qy 782 IPTVGMAGVILLTLVIGIILAVVFIIRIKKDKGNVESQAQASEVKPLRS 832  
Db 781 IPTVGMAGVILLTLVIGIILAVVFIIRIKKDKGNVESQAQASEVKPLRS 827

RESULT 3  
I46536  
Ksp-cadherin - rabbit  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C:Date: 14-Feb-1997 #sequence\_revision 14-Feb-1997 #text\_change 05-Nov-1999  
C:Accession: I46536  
R:Thomson, R.B.; Igarashi, P.; Biemesderfer, D.; Kim, R.; Abu-Alfa, A.; Soleimani, M.  
J. Biol. Chem. 270, 17594-17601, 1995  
A:Title: Isolation and cDNA cloning of Ksp-cadherin, a novel kidney-specific member o  
A:Reference number: I46536; MUID:95340560; PMID:7615566  
A:Accession: I46536  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-829 <THO>  
A:Cross-references: EMBL:U28945; NID:g902885; PIDN:AAC48472.1; PID:g902886

Query Match 22.1%; Score 954; DB 2; Length 829;  
Best Local Similarity 31.6%; Pred. No. 5.7e-52;  
Matches 251; Conservative 142; Mismatches 354; Indels 48; Gaps 22;

Qy 62 LTGET----DNIFVIERE-GLLYYNRALDRETRSTHNLQVAALDANGIIVEGVPVITIEV 116  
Db 57 LSGDGVAAEGPFSVEAESGFLLVTRALDREQAQYQIVTLEAEDGHVLMGPQSVTVHV 116

Qy 117 KDINDNRFTFQSKYEGSVQRNRPCKPFLYVNVATDLDPATPNQLYQIIVTOLPMINN 176  
Db 117 KDENDQVQFQSKYEGSVQRNRPCKPFLYVNVATDLDPATPNQLYQIIVTOLPMINN 176

Qy 177 VMYFOINKTKAISLREGSOELNPAKNPSYNLVSVKDMGGOSENFSFDTSDVITIEV 236  
Db 177 PDVFLERLALALSPESGAGFHALGEPYQLLVQVDMGDOA-SGHQATATVEISIVE 235

Qy 237 NIWAKPVEVENVSTDPHPKIKITQVRWNPDAQYSLVDKEKLPFRPFSDQEGDIYVTP 296  
Db 236 STWTPLPVPVHLAENLVYPVPHLAQVHWSGGVDVHRL---ESQPPGPFVDVTEGLKLYVTG 292



Qy 297 PLDREKDAVYFAVAKDEYKGLPSYPLEIHKVVKDINDNPPTCP-----SPVTVEEVOENE 353  
 Db 293 ELDRQAQYVQVQAQNSRGEDYAEPLHLLVVTVDNDHAPVCPGRPPVSPVPELSP- 351  
 Qy 354 RLGNISICTLAHDDRENTANSFLNRYRIVEOTPKL-PMDGLFLIQTAYGMLQAKSLKK 412  
 Db 352 --GTAVTTLAEDADAPGSPNSHVYVLLSPPEQEGEGGAQDPTSGSVSLGAAPLEA 409  
 Qy 413 QDTPOYNLTIEV-----SDKDKFTLCFQVINDINDQIPIPEKSDYGNLTIAEDTNI 465  
 Db 410 GQ-----NMLQVLAVDLAGAEGSLSTCEVAVTVTDVNDHAPEFTSSQVGPVSLPEDTEP 465  
 Qy 466 GSTILTQATDAD-EPTGSGSKIL-YHIIRKDSGRGLVDTPHNTGYVILIKKPLDFET 523  
 Db 466 GTLVATLTATDADLEP---AFRLMDETIEAGDGEFTGLDWEPSDGHVQVLYLLKNLSYEA 522  
 Qy 524 AAVSNIVFKAENPEPLVFGVKNYASSFAKFTLIIVTDVNEAPQFSQHVFOAKVSEDVAIGT 583  
 Db 523 APSHTVVVVVNRNVETV-GPGGPGATATVTLVKEKVMPPRLQKSEYADIPVNAFAGS 581  
 Qy 584 KGVNVTAKDPEGLDISYSRGRDTRGWLKIDHVTGEIFSVAPLD-REAGSPYRVQVAVTEV 642  
 Db 582 FLTLIQAPENPAGALRESLVNDSEGWFCIQKVSSEVHTARELOGARPGDSYTLVLEAQDA 641  
 Qy 643 GGSLSVSEFHLILMDVNDNPPRLAKDYTGLEFCHPLSPAGSLIFATDD-DOHLFRGP 701  
 Db 642 DAPRLSTSAALVIHFLRAPAPALPALPMPSPRHLCTPRQDHVLPAPSEDDPMATGHGP 701  
 Qy 702 HFTFSLG-SGSLONDWEVSKINGTHARLSTRHTDDEERAYVVLIRINDGRRPPLLEGIVSL 760  
 Db 702 -YSFALGPNPTVQRDLQPLNDSHALFTLALHWVEPREHIVPVVVSODAR-----VWQL 755  
 Qy 761 P--VTFSC-VEGSCFRPAGHTQIPTVGMAGVIGLLTILVIGILAVFTRIRKDKGD 817  
 Db 756 PVRVVVCRNTEGECMRKVKRMKCMPTKLSAVGLVGTFLAAGFFLILIFTHLARKKD 815  
 Qy 818 NVESAQAASEVKKPLRS 832  
 Db 816 --LDAPADNV-PLKA 827

RESULT 4  
 LUCHCR  
 R-cadherin precursor - chicken  
 C;Species: Gallus gallus (chicken)  
 C;Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 16-Jun-2000  
 C;Accession: JH0424  
 R;Inuzuka, H.; Miyatani, S.; Takeichi, M.  
 Neuron 7, 69-79, 1991  
 A;Title: R-cadherin: a novel Ca2+-dependent cell-cell adhesion molecule expressed in the  
 A;Reference number: JH0424; MUID:91299341; PMID:1712604  
 A;Accession: JH0424  
 A;Molecule type: mRNA  
 A;Residues: 1-913 <INU>  
 A;Cross-references: GB:D14459; GB:D00849; NID:Q222854; PIDN:BAA03356.1; PID:Q222855  
 A;Experimental source: retina  
 C;Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to  
 C;Superfamily: cadherin; cadherin repeat homology  
 C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; retina; transmembr  
 F;1-26/Domain: signal sequence #status predicted <SIG>  
 F;27-166/Domain: propeptide #status predicted <PRO>  
 F;167-913/Product: R-cadherin #status predicted <NAT>  
 F;167-913/Domain: extracellular #status predicted <EXT>  
 F;169-274/Domain: cadherin repeat homology <CR1>  
 F;244-249/Region: cadherin binding #status predicted  
 F;277-389/Domain: cadherin repeat homology <CR2>  
 F;392-504/Domain: cadherin repeat homology <CR3>  
 F;507-612/Domain: cadherin repeat homology <CR4>  
 F;613-721/Domain: cadherin repeat homology <CR5>  
 F;722-753/Domain: transmembrane #status predicted <TM>  
 F;754-913/Domain: intracellular #status predicted <INT>  
 F;870-885/Region: serine-rich  
 F;280,409,554,629,699/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 16.5%; Score 715; DB 1; Length 913;  
 Best Local Similarity 26.9%; Pred. No. 7.2e-37;  
 Matches 203; Conservative 117; Mismatches 329; Indels 106; Gaps 19;

Qy 123 RPTFLQSKYEGSVQRNSRQPKPLFYV-----NATDLDPATPNQQLYYQIV 169  
 Db 31 KPGFSEEDYTAFAVSINIMEGOKLLKVKFNCCAGNKGVRYETNSLDFKVRADGTMY--AVH 88  
 Qy 170 QLPMINVMYFOI-----NNKTGATSLREGSOE 198  
 Db 89 QVQMASQQLILMVTAMDPOFLGRWEAIVRFLVGEKLOHNGHKPKRSGPVDLAQQSDT 148  
 Qy 199 LNPAKNPSNLVIVSKDMGGGSENSFSDTTSVDIIVTENIWKAPKPVEMVENSDDPIK 258  
 Db 149 LLPWRHQ-----SAKGLRQKRD-----WVIP-PINVPENSRGPPQ 186  
 Qy 259 ITQVRN---DPGAQYSL--VDKEKLPFRPFSIDQ-EGDIYVTOPLDREKDAVYFAVA 312  
 Db 187 LVIRSKDKDEIHIRYSITGVGADQPPMEVESIDPVSGRMVYVTRPMDEERASYHLRAHA 246  
 Qy 313 KDEYKGLPSYPLEIHKVVKDINDNPPTCPSPVTVEVOENRNLNSIGTLTAHDOENT 372  
 Db 247 VDMNGKNVENPIDLIYIVDMNDNRPEFVINGVDSDEGSKPGTYVMTVTANDADSTT 306  
 Qy 373 ANSFNTYRIVEQTPKPLMDGLFLIQTAYGMLQAKSLKKQDTPOYNLTIEVSKD---- 428  
 Db 307 ANGMVRYRIVQTQSPSQNMFTINSETGDIVTVAAGLDREKVVQVWVIVQATMEGNLN 366  
 Qy 429 --FYLTFVQINVIDINDQIPIPEKSDYGNLTIAEDTNIIGTILTIQATDADEFTGSSK 486  
 Db 367 YGLSNTATATITVDVNDNPEFTTSTYSG--EVPENRVEVVVANLTVMDBRDQPHSPNMN 424  
 Qy 487 ILYHIIRKDSGRGLVDTPHNTGYVILIKKPLDFEFAAVSNIVFKAENPEPLVFGVKYN 546  
 Db 425 AIYRIISGDSGHTFIRTPVTNEGVMVYKAVDYENRRAFLVWVSNQAPLASGIMS 484  
 Qy 547 ASSFAKFTLIIVTDVNEAPQFSQHVFOAKVSEDVAIGTKVGNVTAKDPEGL---DISYSLR 603  
 Db 485 FQSTAGVTISVTDVNEAPYFPTNHLIRLEGGVPTGTVLTTFSAVDPDRFQQAQVRSKL 544  
 Qy 604 GDRGWLKIDHVTGEIFSVAPLDREA---GSPYRVQVAVTEVGGSSLSVSEFHLILMD 659  
 Db 545 SDPANLNLINATNGOITTAADVLDRESDYIKNNVYEATFLAADNGIPASGTTQIYLIID 604  
 Qy 660 VNDNPPRLAKDYTGLEFCHPLSPAGSLIFEAT---DDOHLFRGPHFTFSLGS--GSLQ 713  
 Db 605 INDNAPELLPEAQI-----CEKPNLNVINTAADADIDPNV--GP-FVFEPSVPSAVR 656  
 Qy 714 NDWEVSKINGTHARLSTRHTDDEERAYVVLIRINDGRRPPLLEGIVSLPVTFCSCVEGSCF 773  
 Db 657 KNWITRILNGDYAQLSLRIMYLEAGVYDVPIIVTDSGNPPLYNTSIKVKVCPDENGDC 716  
 Qy 774 RPAGHGQIPTVGMAGVIGLLTTLVIGILAVFI 808  
 Db 717 TTIG---AVAAAGLGTGAITAILICIITLTMVLL 748

RESULT 5  
 A47543  
 R-cadherin precursor - mouse  
 C;Species: Mus musculus (house mouse)  
 C;Date: 03-May-1994 #sequence\_revision 26-May-1994 #text\_change 22-Jun-1999  
 C;Accession: A47543; I55501  
 R;Hutton, J.C.; Christofori, G.; Chi, W.Y.; Edman, U.; Guest, P.C.; Hanahan, D.; Kell  
 Mol. Endocrinol. 7, 1151-1160, 1993  
 A;Title: Molecular cloning of mouse pancreatic islet R-cadherin: differential express  
 A;Reference number: A47543; MUID:94067164; PMID:8247017  
 A;Accession: A47543  
 A;Molecule type: mRNA  
 A;Residues: 1-913 <HUT>  
 A;Cross-references: EMBL:X69966; NID:9429111; PIDN:CAA49589.1; PID:9429112  
 R;Matsunami, H.; Miyatani, S.; Inoue, T.; Copeland, N.; Gilbert, D.; Jenkins, N.; Tak

IJHUCN  
cadherin 2 precursor - human  
N:Alternate names: N-cadherin; neuronal cadherin  
C:Species: Homo sapiens (man)  
C:Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 22-Jun-1999  
C:Accession: A38870; S11487; JQ0751; S13799  
R:Reid, R.A.  
submitted to the EMBL Data Library, November 1990  
A:Reference number: A38870  
A:Accession: A38870  
A:Molecule type: mRNA  
A:Residues: 1-906 <REI>  
A:Cross-references: EMBL:X54315; NID:g34998; PIDN:CAA38213.1; PID:g34999  
R:Reid, R.A.; Hemperly, J.J.  
Nucleic Acids Res. 18, 5896, 1990  
A:Title: Human N-cadherin: nucleotide and deduced amino acid sequence.  
A:Reference number: S11487; MUID:91016946; PMID:2216790  
A:Accession: S11487  
A:Molecule type: mRNA  
A:Residues: 1-340, 'N', 342-698, 'R', 700-704, 'F', 706-906 <RE2>  
A:Cross-references: EMBL:X54315  
A:Note: this sequence has been revised in reference A38870  
F:Walsh, F.S.; Barton, C.H.; Putt, W.; Moore, S.E.; Kelsell, D.; Spurr, N.; Goodfellow  
J. Neurochem. 55, 805-812, 1990  
A:Title: N-cadherin gene maps to human chromosome 18 and is not linked to the E-cadherin  
A:Reference number: JQ0751; MUID:90347462; PMID:2384753  
A:Accession: JQ0751  
A:Molecule type: mRNA  
A:Residues: 160-194, 'IR', 197-211, 'L', 213-227, 'Q', 229, 'N', 231-235, 'G', 237-248, 'T', 250-  
A:Cross-references: GB:M34064

[illegible]

```
Db 257 VIDMNDNPEFLHVVNGTVPESKPGTYVMTVAIDADDNALGMLRYRIVSQAPSTP 316
Qy 390 MDGLFLIQTAYAGMLQAKQSLKQDTPQYNLTIEVSKD-----FKTLCFQVQINVIDIN 443
Db 317 SPNFTINNETGDIITVAAGLQREKVKQYTLIIQATDEGPTVGLSTATAVITVDVN 376
Qy 444 DOIPIF-EKSDYGNLTAEADNIGSTILTIQATDADEFTGSSKILYHIIKRGDSEGRIGV 502
Db 377 DNPPEFTAMTFYGEV---PENRVDIIVANLTVTDKDPHTPAWNAVYRISGDDPTGRFAI 433
Qy 503 DTDPTNFTGYVLIKPLDFEATAVSNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNE 562
Db 434 QTDPSNSDGLTVVVKPIDFETNRMFVLTAAENQVPLAKGQHPQSTATVSVTVIDVNE 493
Qy 563 APQFSQHVFAKVSDDVAIGTKVGNVTAQDEPL-----DISYSLRGDTRGWLKIDHVTGEI 619
Db 494 NPYFAPNPKIIRQEBGLHAGTMTLFTTAQDPRYMQQIRYTKLSDPANWLKIDPVNGQI 553
Qy 620 FSVAPLDREA-----GSPYRVQVQVATEVGGSSLSVSEFHLILMDVNDNPPRLAKDYTGIF 675
Db 554 TTIADVLDRESPNVKNINYNATFLASDNGIPPMMSGTGLQIYLLDINDNAPQVL----- 606
Qy 676 FCHPLSA-----PGSLIPFATDDQHLFRGPH-FTFSLGSSGLONDEWVSKNGTHAR 727
Db 607 ---PQAEETCETPDPSNINITALDYDIDPNAGPFAFDLPLSPVTIKRNTITRLNGDFAQ 663
Qy 728 LSTRHTDPEERAYVLLIRINDGGRPLEGIVSLPVTFCSC-VEGSCFRPAGHQGTIPV- 785
Db 664 LNLKIKFLEAGIYEVPIITDSGNPPKSNISILRVKVCQDCSNGDC-----TDVDRIV 716
Qy 786 --GMAVGILLTLLVIGIILAVFIRIKKDKGNVESQAQASEVKP 829
Db 717 GAGLGTGAIALLCIIILLILVLFVVMKRRDKERQAKQLLIDP 762
```

## RESULT 7

IUCHCN

N-cadherin precursor, neuronal - chicken

C:Species: Gallus gallus (chicken)

C:Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 22-Jun-1999

C:Accession: A29964

R:Hatta, K.; Nose, A.; Nagafuchi, A.; Takeichi, M.

J. Cell Biol. 106, 873-881, 1988

A:Title: Cloning and expression of cDNA encoding a neural calcium-dependent cell adhesion

A:Reference number: A29964; MUID:88153917; PMID:2831236

A:Accession: A29964

A:Molecule type: mRNA

A:Residues: 1-912 &lt;HAT&gt;

A:CROSS-references: GB:X07277; NID:g63649; PIDN:CAA30258.1; PID:g63650

C:Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to

C:Superfamily: cadherin; cadherin repeat homology

C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane pro

F:1-28/Domain: signal sequence #status predicted &lt;SIG&gt;

F:29-164/Domain: propeptide #status predicted &lt;PRO&gt;

F:165-912/Product: N-cadherin #status predicted &lt;MAT&gt;

F:165-720/Domain: extracellular #status predicted &lt;EXT&gt;

F:167-272/Domain: cadherin repeat homology &lt;CR1&gt;

F:242-247/Region: cadherin binding #status predicted

F:275-387/Domain: cadherin repeat homology &lt;CR2&gt;

F:390-502/Domain: cadherin repeat homology &lt;CR3&gt;

F:505-611/Domain: cadherin repeat homology &lt;CR4&gt;

F:612-720/Domain: cadherin repeat homology &lt;CR5&gt;

F:721-752/Domain: transmembrane #status predicted &lt;TM&gt;

F:753-912/Domain: intracellular #status predicted &lt;INT&gt;

F:869-884/Region: serine-rich

F:278,330,407,578,628,657/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 16.18; Score 696; DB 1; Length 912;

Best Local Similarity 26.98; Pred. No. 1.le-35;

Matches 209; Conservative 117; Mismatches 328; Indels 124; Gaps 22;

Qy 126 FLQSKYEGSVQRNSRPGKPFLLVNVATDLDPPATPNGLYQIVIQLPINNVMFQIN-- 183

```
Db 41 FPEOVHSAVVSRSVHGGOPLLNVRFSQD-----NRKIYFGSEP 81
Qy 184 -----NKTGALSRLREGSQELNPAKNPSYNLVISVKMGOS-----ENFSFDTT 228
Db 82 EDFRVGEDGVVYAESFOLSAPTE-----FVVSARDKETQEEQMKVKLPEPAFTGAS 136
Qy 229 SVDIIVTENI-----WKAPKPVEMVENSTDPHPKIKITQVRWMDPGA 269
Db 137 EKDKKIEDIIFPMQYKDSHLLKQKRDWIP-PINLPENSRRGPPQELVRIR----- 189
Qy 270 QYSLVDKEKLPFRFPESI-----DQ-----EGDIYVTPQDLREKDAYVYAVAK 313
Db 190 ----SRRDKSLSLRYSVTGPAGDQPTGIFINPISGSLSVTKPLDREQIAFHLRAHAV 245
Qy 314 DEYGKPLSYPLEIHKVKVDINDNPPTCPSPVTFVEVQNERLGNISGTLTJAHDRDEENTA 373
Db 246 DVNGQOVENPDIDVINMDMNDNRPFLHQVWNGTVPEGSRPGTYVMTVAIDADDNPAQ 305
Qy 374 NSFLNRYIVEQTPKLPMDGLFLIQTAYAGMLQAKQSLKQDTPQYNLTIEVSKD----- 428
Db 306 NGMLRYRILSQAPSSPSNMFTINNETGDIITVAAGLDREKVQYVQYTLIIQATDMEGNPTY 365
Qy 429 -FKTLCFQVQINVIDINDQIPIF-EKSDYGNLTAEADNIGSTILTIQATDADEPFGSSK 486
Db 366 GLSNTATAVITVDVNDNPPPEFTAMTFYGEV---PENRVDIVANLTVTDKDPQHTPAWN 422
Qy 487 ILYHIKSGSGLRGVDPDHTNTGYVLIKKPLDPEATAVSNIVFKAENPEPLVFGVKYN 546
Db 423 ARYQMTGGDGTQGTFTLIDPNNSDGLVTVKPIDETNRMFVLTVAENQVPLAKGIQHP 482
Qy 547 ASSFAKFTLIVTDVNEAPQFSQHVFAKVSDDVAIGTKVGNVTAQDP-----EGLDISYSL 602
Db 483 PQSTATVSIITVDYNESFYFVNPKNLVRQEEGLAGSLMTFTTARDPDRYMQOTSLSYSK 542
Qy 603 RGDTRGWLKIDHVTGEIYSVAPLDREA-----GSPYRVQVQVATEVGGSSLSVSEFHLILM 658
Db 543 LSDPANWLKIDPVNGQITTTAVLDRESIYVQNNMYNATFLASDNGIPPMMSGTGLQIYLL 602
Qy 659 DVNDNPPRL-AKDYTGLEFFCHPLSAPGSLIPEATDDQHLFRGPHFTFSLGSG--SLQND 715
Db 603 DINDNAPQVNPKEAT---TCETLQ-PNAINTAVDPDIDPNAGP-FAFELPDSPPSIKRN 657
Qy 716 WEVSKINGTHARLSTRHDFDEERAYVVLIRINDGGRPLEGIVSLPVTFCSC-VEGSCFR 774
Db 658 WTIVRISGDHAQLSLRIRFLEAGIYDVIVITDSGNPHASSTSVLKVVCQCCDINGDC-- 715
Qy 775 PAGHOTGIPTV---GMAVGILLTLLVIGIILAVFIRIKKDKGNVESQAQASEVKP 829
Db 716 ----TDVDRIVGAGLGTGAIALLCIIILLILVLFVVMKRRDKERQAKQLLIDP 768
```

## RESULT 8

JUMSCN

N-cadherin precursor, neuronal - mouse

C:Species: Mus musculus (house mouse)

C:Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 22-Jun-1999

C:Accession: A32759; A46163

R:Myatani, S.; Shimamura, K.; Hatta, M.; Nagafuchi, A.; Nose, A.; Matsunaga, M.; Hat

Science 245, 631-635, 1989

A:Title: Neural cadherin: role in selective cell-cell adhesion.

A:Reference number: A32759; MUID:89346748; PMID:2762814

A:Accession: A32759

A:Molecule type: mRNA

A:Residues: 1-906 &lt;MIY&gt;

A:CROSS-references: GB:M31131; NID:g192327; PIDN:AAA37353.1; PID:g309125

R:Myatani, S.; Copeland, N.G.; Gilbert, D.J.; Jenkins, N.A.; Takeichi, M.

Proc. Natl. Acad. Sci. U.S.A. 89, 8443-8447, 1992

A:Title: Genomic structure and chromosomal mapping of the mouse N-cadherin gene.

A:Reference number: A46163; MUID:92409532; PMID:1528849

A:Accession: A46163

A:Status: preliminary

A:Molecule type: DNA

A; Residues: 839-906 <M2>  
A; Cross-references: GB:S45011; NID:g256010; PIDN:AAB23356.1.; PID:g256011  
A; Note: sequence extracted from NCBI backbone (NCBIN:I13759, NCBIP:I13760)  
C; Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought t  
C; Superfamily: cadherin; cadherin repeat homology  
C; Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane pro

F; 1-27/Domain: signal sequence #status predicted <SIG>  
F; 28-159/Domain: propeptide #status predicted <PRO>  
F; 160-906/Product: N-cadherin #status predicted <MAT>  
F; 160-714/Domain: extracellular #status predicted <EXT>  
F; 162-267/Domain: cadherin repeat homology <CRI>  
F; 237-242/Region: cadherin binding #status predicted  
F; 270-382/Domain: cadherin repeat homology <CR2>  
F; 385-497/Domain: cadherin repeat homology <CR3>  
F; 500-605/Domain: cadherin repeat homology <CR4>  
F; 606-714/Domain: cadherin repeat homology <CR5>  
F; 715-746/Domain: transmembrane #status predicted <TMN>  
F; 747-906/Domain: intracellular #status predicted <INT>  
F; 865-878/Region: serine-rich  
F; 190,273,325,402,572,651,692/Binding site: carbohydrate (Asn) (covalent) #status predic

Query Match 16.1%; Score 694.5; DB 1; Length 906;  
Best Local Similarity 27.6%; Pred. No. 1.4e-35;  
Matches 211; Conservative 122; Mismatches 332; Indels 99; Gaps 23;

QY 117 KOINDNRPTFLQSKYEGSVQRNGRPGKFFLYVNATDLDPATPNGLYYOIVQLPMMNN 176  
|::: | | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 47 KDVHEGQP-LLNKFSNCNRK-----RKVYESPADFKVEDGTGY--AVRSFPLTAE 98  
  
QY 177 VMVFQI---NNKTG-----AISLTREGSQBELNPAKNP-SYNLVISVKDM---GGQSENSF 224  
| | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 99 QAKFLIYAQDKTEQEKQWQVAVNLISREPTLTETPEMKDEHEIEIVFPROLAKHGALORQK 158  
  
QY 225 SDTTSVDIIVTENIWKAPKVEMVENSTDPHPKITQVRWN-----DPGAQYS 272  
| - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |  
Db 159 RD-----WVIP-PINLPENSGRPPQELVRIIRSROKNLSRLSYVTGPGA--- 202  
  
QY 273 LVDEKEILPRPFPSIDQ--EGDIYVTPQDLDEKRAYFYAVAKDEYGKPLSLPLEIHVVKK 331  
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 203 ----DQPTGFIFINPISSQLSVTKPLDRILLAFHRAHAVDINGNQVENPIDIVNVI 258  
  
QY 332 DINDNPTCTSPVTVFVSQNERLGSIGTLTAHDREENTANSFLAYRVVEOPKPKMD 391  
| : | | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 259 DMNDNRPEFLHQVWNGSVPEGSKPTGVMTVTAIDADPNALNGMLRYIRLSQAPTSP 318  
  
QY 392 GLFLIQTYAGLMQLAKOSLKQQDTPQYNLTATEVSDKD-----FKTLCFVQINVINDNQ 445  
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 319 NMFTINNETGDIIITVAAGLDREKVQOYLIIQATDMEGNPYGLSNTATAVITVDNDN 378  
  
QY 446 IPIF-EKSDXGNLTFLAEDNTIGSTILTQATDADEPTGSKILIYHIKGDSERGLGVD 504  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 379 PPEFTAMTFYGEV---PENRVDIVANLTVDKQHTPAWNAARYISGGDPTGRFAILT 435  
  
QY 505 DPHTNTGVYIKKPLDFETRAVSNIVFAENPEPLVFGVKYNASSPAKFLIIVTDVNEAP 564  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 436 DPNSENDGLVTVKPDIDEFTRMFVLTVAAENQNVPFLAKGIQHPPQSTATSVTVIVDNEP 495  
  
QY 565 QFSOHVFQAQVSDVAIGTKVGNVTAKDQEGL---DISYSLRGDTGRWLKIDHVTGEFS 621  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 496 YFAPNPKLIHQEGLHAGTMLTTIQAQDDPRYMQONIRYTKLSDPANWLKIDPVNGQLTT 555  
  
QY 622 VAPLDREA---GSPYRVQVVADEVGSSISEFHLLIMDVNDPPRLAKDYTGFLFCF 677  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 556 IAVLDRESPYQNNIYNATFLASDNGIPPHSGTGLTIQLYLLDINDNAPQVL----- 606  
  
QY 678 HPLSA-----PGSLIFEATDDQHLFRGPH-FTFSLGSGSLQNDWEVSKINGTHARLS 729  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 607 -PQEARETCETPEPNSINIAALDYIDFNAGPFDLPLSPVTIKRNWTINRLNGDFAQLN 665  
  
QY 730 TRHTDFERAYVVLIRINDGCRPPLEGIVSLPVTFCS-VEGSCFRPAGHOTIPTV--- 785  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 666 LKIKLEAGIYEVPDIITDSGNPPKSNISILRVKVCQDSNGDC-----TDVDRIVGA 718

**QY**   786 GMAVGILLTLLVIGIILAVVFIRIKKDKGNVESQAASEVKP 829  
               | : | : ||||| :| :| :| :| :| :  
**Dd**   719 GLGTGAIIALLCIIILIILLVFMFVMWKKRDRKAQLLIDP 762  
               | : | : ||||| :| :| :| :| :| :

**RESULT 9**

LJB0CN  
N-cadherin precursor - bovine (fragment)  
C:Species: Bos primigenius taurus (cattle)  
C>Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 22-Jun-1999  
C:Accession: S11693  
R:Lilaw, C.W.; Cannon, C.; Power, M.D.; Kiboneka, P.K.; Rubin, L.L.  
EMBO J. 9, 2701-2708, 1990  
A>Title: Identification and cloning of two species of cadherins in bovine endothelial  
A:Reference number: S11693; PMID:90360979; PMID:2390969  
A:Accession: S11693  
A:Molecule type: mRNA  
A:Residues: 1-877 <LIA>  
A:CROSS-references: EMBL:X53615; NID:g164; PIDN:CAA37677.1; PID:g664894  
C:Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are though  
C:Superfamily: cadherin; cadherin repeat homology  
C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane  
F:1-130/Domain: propeptide (fragment) #status predicted <PRO>  
F:131-877/Product: N-cadherin #status predicted <MAT>  
F:131-885/Domains: extracellular #status predicted <EXT>  
F:133-238/Domains: cadherin repeat homology <CRI>  
F:208-213/Region: cadherin binding #status predicted  
F:241-353/Domains: cadherin repeat homology <CR2>  
F:356-468/Domains: cadherin repeat homology <CR3>  
F:471-576/Domains: cadherin repeat homology <CR4>  
F:577-685/Domains: cadherin repeat homology <CR5>  
F:686-717/Domains: transmembrane #status predicted <TMN>  
F:718-877/Domains: intracellular #status predicted <INT>  
F:836-849/Region: serine-rich  
F:161,244,296,373,543,593,622,663/Binding site: carbohydrate (Asn) (covalent) #status

Query Match      15 q%    Score 688.5; DB 1; Length 877;  
Best Local Similarity    26.7%; Pred.No. 3.le-35;  
Matches    203; Conservative 126; Mismatches 343; Indels 87; Gaps 19;

**Qy**   126 FLOSKYEGSVQRNSRGPKFLYNATLDPPATNGOLYTQIVIQLP-----MINN 176  
               | : | : ||||| :| :| :| :| :| :  
**Dd**   7 FPEDYSANLSRDVLGEPLNLNVFSNC-----NGRRKVYESPADFKVDGDGMVA 60  
               | : | : ||||| :| :| :| :| :| :

**Qy**   177 VWFYOINKTKGAISLTREGSGOELPANPNPSNYLVISKDMGGQSENESFDTSVDIIV-- 234  
               | : | : ||||| :| :| :| :| :| :  
**Dd**   61 VRSPPLSSEHKSFIYAQDKE----TOEKQQWAVALSLKPALPEDSVKSESREIEEIVFP 115  
               | : | : ||||| :| :| :| :| :| :

**Qy**   235 -----TENWKAKPVMEVNSTDPHPKITOVERN-----DGCAQ 270  
               | : | : ||||| :| :| :| :| :| :  
**Dd**   116 RQVTKHNGYLQORQKDWWIP-PINLPENSRGFPOELVRHSRKDNLSLRYSVTGPCEA- 173  
               | : | : ||||| :| :| :| :| :| :  
**Qy**   271 YSLVDKEKLPRPFSDIO-EGLDIVTPDLDEEKDAVVFAVADKEYCKPLSYPLEIHVK 329  
               | : | : ||||| :| :| :| :| :| :  
**Dd**   174 ----DOPTGCIFIIINPISQLSWTRPLDRELTAHFHLRAHVADINGNQVENPIDVIN 227  
               | : | : ||||| :| :| :| :| :| :

**Qy**   330 VKDINDNPPTCPSPVTVFEQENERLGINSIGTLTAHDRDEENTANSFLNYRIVEQT PKLP 389  
               | : | : ||||| :| :| :| :| :| :  
**Dd**   228 VIDNDNRPELFHQWNCTVPTEGSKGTYMTVTAIDADDNALMGMLRYRLSQAPSTP 287  
               | : | : ||||| :| :| :| :| :| :

**Qy**   390 MDGLFLTIQTAGMILQLAQSUKKODTFOYNLTUIEVSKD-----FKTCLCFVQINVVIDN 443  
               | : | : ||||| :| :| :| :| :| :  
**Dd**   288 SPNNFTINNNTGITITAAAGLDREKVOQYTLIIQATDMEGNPTYGLSNATAFATVTTDVN 347  
               | : | : ||||| :| :| :| :| :| :

**Qy**   444 DQEIF-EKSDYGNIETAEADTNIGSTITLTQATDADEPGTGSSKIIXHI IKGDSEGRLGV 502  
               | : | : ||||| :| :| :| :| :| :  
**Dd**   348 DNPEFTAMTFYEY---PENRVDTVIANLVTDKQPHTPAWNAIYRISCGDSPAGRFAI 404  
               | : | : ||||| :| :| :| :| :| :

**Qy**   503 TDPTHNTGYVIHKPKPDFTAASVISNFKAENSEPELVFGKYNASFFAKEFTLIIVTDVNE 562  
               | : | : ||||| :| :| :| :| :| :  
**Dd**   405 QTDPSNDGLVTYVKPIDFETRMYVLTVAENQVPLAKGIQHPPQSTATVSVTVIDVNE 464  
               | : | : ||||| :| :| :| :| :| :

**Qy**   563 APQFSQHVFQAKVEDSAIGTKVGNVTA KDPEGL---DISVSLGRDRTGWLUKDIHWTCGEI 619  
               | : | : ||||| :| :| :| :| :| :

```
Db 465 NPYFAPNPKIIRKQEGHAGTTLTTFTADDPDRYMQQNIIRYTKLSDPANWLKIDSVNGOI 524
Qy 620 FSVAPLDREA-----GSPYRVQVAVTEVGGSSLSVSEFHLILMDVNDNPPRLAKDYTGIF 675
Db 525 TTIIVLDRESNPVKANIYNATFLASDNGIPKMSGTGTQIYLLDINDNAPQVLPOEAEI- 583
Qy 676 FCHPLSAPGSLIFEATDDQHLFRGPH-FTFSLGSGSLQNDWEVSKINGTHARLSRTHD 734
Db 584 -CE-TPDPNSINITALDYIDNPAGPFAFDLPSPVTKRNTITRLNGDFAQLNKKIKF 641
Qy 735 FEERAYVVLIRINDGGRPLEGIVSLPTFCSC-VEGSCFRPAGHOTGIPV---GMAVG 790
Db 642 LEAGIYEPIIITDSGNPKPNISILRVKVCODSNGDC-----TDVDRIVGAGLGTG 694
Qy 791 ILLTTLTLLVIGIILAVVIFIRIKKDKGNVESQAQSEVKP 829
Db 695 AIIAILLCIIILLIILVFMVVMKRRDKERQAKQLLIDP 733

RESULT 10
C38992
cadherin 4 precursor - human
N:Alternate names: R-cadherin
C:Species: Homo sapiens (man)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 20-Aug-1999
C:Accession: C38992
R:Suzuki, S.; Sano, K.; Tanihara, H.
Cell Regul. 2, 261-270, 1991
A:Title: Diversity of the cadherin family: evidence for eight new cadherins in nervous t
A:Reference number: S24305; MUID:91283540; PMID:2059658
A:Accession: C38992
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-916 <SUZ>
A:Cross-references: GB:L34059; NID:g506409; PIDN:AAA35627.1; PID:g506410
C:Genetics:
A:Gene: GDB:C8H4
A:Cross-references: GDB:622850
A:Map position: 16q24.1-16qter
C:Superfamily: cadherin; cadherin repeat homology
C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane pro
F:1-27/Domain: signal sequence #status predicted <Sig>
F:28-166/Domain: propeptide #status predicted <PRO>
F:167-913/Product: R-cadherin #status predicted <MAT>
F:167-721/Domain: extracellular #status predicted <EXT>
F:169-274/Domain: cadherin repeat homology <CR1>
F:244-249/Region: cadherin binding #status predicted
F:277-389/Domain: cadherin repeat homology <CR2>
F:300-304/Domain: calcium binding #status predicted <CAB>
F:392-504/Domain: cadherin repeat homology <CR3>
F:507-612/Domain: cadherin repeat homology <CR4>
F:613-721/Domain: cadherin repeat homology <CR5>
F:722-753/Domain: transmembrane #status predicted <TM>
F:754-913/Domain: intracellular #status predicted <INT>
F:870-885/Region: serine-rich
F:283,412,557,632,661,702/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 15.88; Score 684.5; DB 2; Length 916;
Best Local Similarity 29.38; Pred. No. 6e-35;
Matches 176; Conservative 106; Mismatches 288; Indels 31; Gaps 12;

Qy 239 WKAKPKVMENSTDPHPKIKITQVRW---NDPGAQYSL--VDKEKLPRFPESID-QEGDI 292
Db 171 PINVPNSRPPQQLVRIRSDKNDIPIRYSITGVGADOPPMEVFSINSMGRM 229
Qy 293 YVTOPLDREKDAYVYAVAKDEYKPLSYPLEIHVKVDINDNPPTCPSPVTFVEQEN 352
Db 230 YVTRPMDREERASHLRAHAYDMGNKVENPIDLIYIYIDMNDNHPFQINQVYNCSDVEG 289
Qy 353 ERLGNSIGTLTAHDEENTANSLNRYIVETQPKLPMGDLFLIOTYAGMLQAKQSLKK 412
Db 290 SKPGTYVNTIITANDADOSTTANGVMRYRIVTQTQPSQNMFTINSETGDIIVTVAAGMDR 349
```

```
Qy 413 QDTPOYNLTIEVSKD-----FKTCLFVQINVIDINQIPIFEKSDYGNLTLAEDTNIG 466
Db 350 EKVOQYTVIVQATDMEGNLNGLSNATATAITVTDVNDNPNSEFTASTAG--EVPENSVE 407
Qy 467 STILTIQATDADEPTGSSKILYHIKGDSEGRGLGVDTPHTNTGYVYIKKPLDPETAIV 526
Db 408 TVVANLTYMDRDQPHSPNNNAVYRIISGDPGHFSVRTDPVTNEGVTWVKAVDYELNRA 467
Qy 527 SNIVFKAENPEPLVFGVKNASSPAKFTLIVTDVNEAQFESOHVFOAKVSEDVAIGTKVG 586
Db 468 FMLTVWGNQAPLASSGIOMSFQSTAGVTISIMDINEAYFFPSNHLKIRLEEVPPTVLT 527
Qy 587 NVYAKDPEGL---DISYSLRGDTGRGLKIDHVTGIFSVAPLDREA----GSPYRVQVVA 639
Db 528 TFSAVDPDPREMQQAVRYSKLSDPASLHINATNGQITTVAVLDRSLYTKNNVYEATFLA 587
Qy 640 TEVGGSSLSVSEFHLILMDVNDNPPRLAKDYTGFLFFCHPLSAPGSLIFEATDDQHLFR 699
Db 588 ADNGIPPASGTGTLQIYILIDINDNAPPELLKPEAQICERPNLNA---INITAADADVHPNI 644
Qy 700 GPH-PTFSLGSGSLQNDWEVSKINGTHARLSRTHDDEERAYVVLIRINDGGRPLEGIV 758
Db 645 GPYVFELPFVPAAVRKNWTITRLNGDYAQLSLRILYLEAGMYDVPPIIYVDSGNPLSNTS 704
Qy 759 SLPTFTSCVSCGSCFRPAGHOTGIPTVGMVAGILLTLLVIGIILAVV--FIRIKKDKGK 816
Db 705 IIKVKVCPDDNGDCTTIG---AVAAAGLGTGAIVAILICILILLTWMVLLFVMMKREK 761
Qy 817 D 817
Db 762 E 762

RESULT 11
I50116
N-cadherin precursor - zebra fish
C:Species: Brachydanio rerio (zebra fish)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jan-2000
C:Accession: I50116
R:Bitzur, S.; Kam, Z.; Geiger, B.
Dev. Dyn. 201, 121-136, 1994
A:Title: Structure and distribution of N-cadherin in developing zebrafish embryos: mo
A:Reference number: I50116; MUID:9517841; PMID:7873785
A:Accession: I50116
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-783 <BIT>
A:Cross-references: EMBL:X67648; NID:9623209; PIDN:CAA47890.1; PID:9623210
C:Superfamily: cadherin; cadherin repeat homology
F:146-258/Domain: cadherin repeat homology <CDH>

Query Match 15.78; Score 677; DB 2; Length 783;
Best Local Similarity 30.18; Pred. No. 1.4e-34;
Matches 187; Conservative 98; Mismatches 288; Indels 48; Gaps 15;

Qy 239 WKAKPKVMENSTDPHP---IKITQVRWNDPGAQYSLV--DKEKLPRFPESIDQ-EGDI 292
Db 37 WVIP-PVNVLENSRKQFPPELVKIQSDKSNLRYSVTGPCADONPTGLFIIDPISGLL 95
Qy 293 YVTOPLDREKDAYVYAVAKDEYKPLSYPLEIHVKVDINDNPPTCPSPVTFVEQEN 352
Db 96 SVTKPLDREHIPNPHLRAHAVIDINGNQMNPIIDIIINVIDMNDNRPETHQIWNQTVDEG 155
Qy 353 ERLGNSIGTLTAHDEENTANSLNRYIVETQPKLPMGDLFLIOTYAGMLQAKQSLKK 412
Db 156 AKPGTFVMTVTVSQKDDPNTANGMLRYKILSQTPESSSNMFTINNKTKIITVAAGLDR 215
Qy 413 QDTPOYNLTIEVSKD-----FKTCLFVQINVIDINQIPIFEKSDYGNLTLAEDTNIG 466
Db 216 EKVOQYTVIIVQATDMEGNPTGLSNATATAVIRLLDVNDNNAPEFTRETHGEVPEPNVVI 275
Qy 467 STILTIQATDADEPTGSSKILYHIKGDSEGRGLGVDTPHTNTGYVYIKKPLDPETAIV 526
```

Db	276	VTNLTV--TDKEPGTPANNNAVTRIISGDPTRGRFSIPTDPVNBGLVTVVKRPVDFEMNRS	333
Qy	527	SNIVFRAENPEPLVFGVKYNASSFAKFTLLIVTDVNEAPOSQHVQAKVSESDVAIGTKVKG	586
Db	334	FMLTVVADNEVPLASGIHRTROSTATVSIRVIDNESPNFDPNPKQIKLEEGLPQWSMLT	393
Qy	587	NVTA-KDPEGL---DTSYSLRGDTRGWLKIDHVTGBIFSVAPLDREA-----GSPYRVQVW	638
Db	394	TFTAHDDPDYMQOITISYSKLYDPANLLEIDPNNGRISTIAVLDRSPYVKNLNLYATFM	453
Qy	639	ATEVGSSLSVSEFEHLIMDVNDNPRLAKDYTGLGFCHPLSA-----PGSLIFEAT	691
Db	454	ASDNGYPRASGTQLIYLLDINDNAPRV-----FPQEAECVCRPENAINITAV	503
Qy	692	DDQHLPRGRPHFTFSLGS--GSLQNDWEYSKINGTHARLSTRHTDPEERAYVVLTRINDG	749
Db	504	DGDLNPNAGP-YAFELPNRPSDIRRNMWTLTRISGDHAQLSLKISYLESGIYELPISITDS	562
Qy	750	GRPPLGIVSLPVTFFCSC-VEGSCFRPAGHGTPHVGMAVGLLFTLLVIGIILLAVVFI	808
Db	563	GNLPMSTWTTYLRIKVCQCDDHHGDCV---DMERIMAAAGLGTGAITAILICITILLVVLVM	618
Qy	809	RIKKDKGCKDNVESAAQASEVKP	829
Db	619	FVMMKMKRRDKERQAKOLLIDP	639

RESULT 12

B55363  
desmocollin, type 4, short form precursor - human  
C:Species: Homo sapiens (man)  
C:Date: 06-Feb-1995 #sequence\_revision 06-Feb-1995 #text\_change 23-May-1997  
C:Accession: B55363  
E:Kawamura, K.; Watanabe, K.; Suzuki, T.; Yamakawa, T.; Kamiyama, T.; Nakag  
J. Biol. Chem. 269, 26295-26302, 1994  
A:Title: cDNA cloning and expression of a novel human desmocollin.  
A:Reference number: A55363; MUID:95014464; PMID:7929347  
A:Accession: B55363  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-826 <KAW>  
A:Cross-references: GB:D17427  
C:Superfamily: cadherin; cadherin repeat homology  
C:Keywords: cadherin; cadherin repeat homology  
F:138-243/Domain: cadherin repeat homology <CR1>  
F:246-355/Domain: cadherin repeat homology <CR2>

Query Match	15.5%	Score 671.5;	DB 2;	Length 826;
Best Local Similarity	30.8%	Pred. NO. 3.3e-34;		
Matches 187;	Conservative 106;	Mismatches 263;	Indels 51;	Gaps 22;

Qy	239	WKAPKPVEMVENSTDPHPKIKITQVRWNDPGAQVSL-----VDKEKLPREFFFSIDQEG	290
Db	137	W-APIPCMSOENSLGGPPLFQOQE--SDAAQNYTTFYSISGRGVQDEKUNLEIVIERD-TG	193
Qy	291	DIYVQTPLDREEDKDAY--VFYAVAKDEYKGKPSLYPLEIHWVKVDINDNPEPTCPSVTVFE	348
Db	194	NLFCRTPVDREEDVDVFLIAYASTADGYSNDL--PLPLPIRVEDENDNHPVTEALYNFE	251
Qy	349	VOENRIGNSIGTILTAHRODEENTANSFLNYRIVEQTPKLPMDGLFLQTYAGMIQLAKQ	408
Db	252	VLESSRPGTTGVGVCACTDRDEPDTMTRLKYSILOOTPRSP--GLFSVPSTGCVTTTVSH	309
Qy	409	SLKKODTPQYNLATEYSDKDFK-----TLCFVQNLVINDIQIPIFKSDYGNLTLAE	461
Db	310	YLDRVVVDKYSLLMKYQDMQDGFGLIGTSTCI--ITVTDSDNRAPTFQONAYE--AFVE	365
Qy	462	DTNIGSTILTIQATDADEPTGSGSKLYHIKGDSSGRGLVDTPDHTNGYVILIKKPLDF	521
Db	366	ENAFNVILRIPTIEDKDLINANWRNYFTILKNGENGHFKISTDKETNPGVLSVVKPLNY	425
Qy	522	ETAASVNIFFKAPNPEPLVFGV--KYNASSFAKFTLLIVTDNBPAPFSOHVFOAKYSEDA	580

Db	426	EENQVWLEIGVNNEAPPADIPRVTALNRALVTVVHVRDLDGDPCTPAAQYVIKENLA	485
Qy	581	IGTRKGVNTAKDPE--GLDISYSLRGDTRGWLKIDHDVTGEIFSVPALDREAGSP----	633
Db	486	VGSINGYKAYDPENRNGNLRYKKLHPDKGWITIDEISGSIITSKILDREVETPKNELY	545
Qy	634	RQVVATEVGSSLSVSSEPHLLMDVDNPPRLAKDYTGFFCHPLSAPGSLIFEATDD	693
Db	546	NITVALIDKDRDCTGTAVN--IEDNDNPPEIIQOEY-VVICPKMGYTDLI--AVDP	599
Qy	694	DOHLFRGPHTFSLSGSS--LQNDWEVSKINETHARLS-TRHTDFEERAYVVLIRINDGG	750
Db	600	DEPVHGAP-EYFSLPNTSPESILWSLTKVNDTAARLSYQKNAGRFQE-YTIPITVKD--	654
Qy	751	RPLEGIVSLPVTFCSVEGSCRRPGAGHOTGPTVGMNAV-GILLTTLLVIGIILAVVFIR	809
Db	655	RAGOATKLLLRVNLCECTHPTQCRTSRSTGVILCKWAILLAILLGALLFVLLTLVCGV	714
Qy	810	IKKDKGK 816 	
Db	715	FGATKGK 721 	
RESULT 13			
A55363			
C:desmocollin, type 4, long form precursor - human			
C:Species: Homo sapiens (man)			
C>Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 20-Jun-2000			
C:Accession: A55363			
R:Kawamura, K.; Watanabe, K.; Suzuki, T.; Yamakawa, T.; Kamiyama, T.: Nakagawa, Y.: J. Biol. Chem. 269, 26295-26302, 1994			
J>Title: cDNA cloning and expression of a novel human desmocollin.			
A:Reference number: A55363; MUID:95014464; PMID:7929347			
A:Accession: A55363			
A>Status: preliminary; not compared with conceptual translation			
A:Molecule type: mRNA			
A:Residues: 1-896 <RAW>			
A:Cross-references: GB:D17427; NID:9639672; PIDN:BAA04249.1; PID:g685219			
C:Superfamily: cadherin; cadherin repeat homology			
C:Keywords: alternative splicing; transmembrane protein			
F:138-243/Domain: cadherin repeat homology <CR1>			
F:246-355/Domain: cadherin repeat homology <CR2>			
Query Match 15.5%; Score 671.5; DB 2; Length 896;			
Best Local Similarity 30.8%; Pred.No. 3.8e-34;			
Matches 187; Conservative 106; Mismatches 263; Indels 51; Gaps			

Qy	239	WKA	PKPEVMEWENSTDPHPKIKITQVRWMDPCAQYSL-----VDKEKLP	PRPFPSIDQBG	290
Db	137	W-	APICPSMQENSLGPPFLFQQVE--SDAAQNTVFYSISGRGVDEPNLFLYIERD-TG	193	
Qy	291	DIY	VTQPLDREKDAY--VFYAVAKDEYGRPLSPYPLIEHVKKVDINDNP	TPCSPVTVPE	348
Db	194	NLFC	TRPVDRREYDVFLLIAYASTADQYSADL--PULPIRVEDENDNDNPV	TFTEALYNFE	251
Qy	349	VOEN	RLGNSIGTLTAHDREDEENTANSFLNRYVEQTOKPLMPDGLGLTIQY	AGMLQLAQK	408
-	252	VLES	RPGTIVGVVCATDREBDPTMHTRLKXSLIQQTPRSP--GLFSVHP	STGVITVSH	309
Qy	409	SLK	QDTPQYNLTIEVSDKFK-----TLCFVQINVIDINDOIP	IFPKSYGNITLAE	461
Db	310	YLD	REWVKYSLIMKQMDMQDQFGLIGTGTCT--ITVTDSDNDAFT	RQNAYE--AFVE	365
Qy	462	DTN	TGSTILTQATDADEPFTGSKILYHIHKDSEBGRGLGVDTDPHTNTGY	VIKKPLDF	521
Db	366	ENAF	NVEILRIPIEDKOLINTANMRVNFITLKGNGHGFKISDTKETNEG	SVLWVKPLNY	425
Qy	522	ETAA	SVNIVFKAENPEPLVFCV-KYNASSFAKFTLIVTDVNEA	POQSQHVQAKVSEDA	580
Db	426	EENQ	VNLIEIGVNEAPFARDIPKVTALNRLVTVHVRDUDGEPECTP	RAAQYVRKENUA	485
Qy	581	IGTK	VGNVTAKDPE---GLDISYLSRGDTRGWLKIDHVTGEIF	ESVAPLDREAGSP----	633

```
Db 486 VGSKINGKAYDPENRNGNGLRYKKLHDPKGWITIDEISGSIITSKILDREVEPTPKNELY 545
QY 634 RVQVATEVGGSSLSVSEFHLILMDVNDNPPRLAKDYTGLFCHPLSAPGSLIFEATDD 693
Db 546 NIVVLAIDKDRSCTGTFLAVN--IEDVNDNPPPEILQY--VWICKPKMGYTDIL--AVDP 599
QY 694 DOHLFRGPHFTFSLGSGS--LQNDWEYSKINGTHARLS--TRHDFEERAYVVLIRINDGG 750
Db 600 DEPVHAP-FYFSLPNTSPSEIRLSWLTKNVDTAARLSYQKNAGFQ--YTIPTVKD-- 654
QY 751 RPPLEGIVLPVTFCSVEGSCFRPAGHQGTPTVGMV--GILTTLLVIGIILAVVFIIR 809
Db 655 RAGOAAFKLLRVNLCEHTPTQCRATSRSTGTVILGKWMAILAILLGIALFSLVLLTLCVG 714
QY 810 IKKDKG 816
Db 715 FGATKG 721

RESULT 14
IJHJHDB
N;Alternate names: desmosomal glycoprotein III
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 07-May-1999
C;Accession: A40390; S16464
R;Parker, A.E.; Wheeler, G.N.; Arnemann, J.; Pidsley, S.C.; Ataliotis, P.; Thomas, C.L.;
J. Biol. Chem. 266, 10438-10445, 1991
A;Title: Desmosomal glycoproteins II and III. Cadherin-like junctional molecules generat
A;Reference number: A40390; MUID:91244819; PMID:2037591
A;Accession: A40390
A;Molecule type: mRNA
A;Residues: 1-847 <PAR>
A;Cross-references: GB:X56807
A;Note: it is uncertain whether Met-1 is the initiator or whether translation is initia
R;Buxton, R.S.; Cowin, P.; Franke, W.W.; Garrod, D.R.; Green, K.J.; King, I.A.; Koch, P.
J. Cell Biol. 121, 481-483, 1993
A;Title: Nomenclature of the desmosomal cadherins.
A;Reference number: A40392; MUID:93252984; PMID:8486729
A;Contents: annotation; nomenclature
C;Genetics:
A;Gene: GDB:DSC3; DSC2; DSC1; DS
A;Cross-references: GDB:I26552; OMIM:600271
A;Map position: 18q12.1-18q12.1
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: alternative splicing; calcium binding; cell adhesion; duplication; glycopro
F;1-28/Domain: signal sequence #status predicted <SIG>
F;136-847/Product: desmocollin 3b #status predicted <PRO>
F;136-695/Domain: extracellular #status predicted <EXT>
F;138-243/Domain: cadherin repeat homology <CRI>
F;246-355/Domain: cadherin repeat homology <CR2>
F;358-471/Domain: cadherin repeat homology <CR3>
F;474-577/Domain: cadherin repeat homology <CR4>
F;578-680/Domain: transmembrane #status predicted <TM>
F;719-847/Domain: intracellular #status predicted <INT>
F;166.392.546.629/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 15.4%; Score 667.5; DB 1; Length 847;
Best Local Similarity 30.9%; Pred. No. 6.2e-34;
Matches 186; Conservative 104; Mismatches 253; Indels 59; Gaps 21;

QY 239 WKAPKPVEMVENSTDPHPKIKITQVRNDPQAQYSL-----VDKEKLPFRFPFSIDQE- 289
Db 137 W-APIPCSMLENSLGPPLPQQVQ--SDTAQNTIYVIRSGPVQDE--PRNLFYVERDT 192
QY 290 GDIVTPTPLDREKDAYVFAVFAKDEYKGLSPLEIHRVVKDINDNPPPTCPSPVTVFEV 349
Db 193 GNLVCRPVDREQYESFEIATFATPDGYTPPELPLIILKIENDNYPFTETTYFTI 252
QY 350 QENRGNSTGTTTAHRDRDENTANSFLNRYIVEQTPLKPMGDLFLIQTIVAGMLQAKQS 409
```

```
Db 253 FENCRVCTTGVQVCATKDEPDTMHTRLKYSIIQVQ--PSPTLFSMHPTTGVITTTSSQ 310
QY 410 LKKQDTFOYNLTIEVSKD-----FKTLCFQVINVIDINOIPIFEKSDYGNLTLAEDTN 464
Db 311 LDRELIDKYOLKIKVQMDGQYFGLQTTSTCIINIDVDNHLPTFTTSY--VTSVEENT 368
QY 465 IGSFILTIQATDADPEPTGSSKILYHIKGDSEGRGLGVDTPHTNTGVVIKKPLDPE-- 522
Db 369 VDVEILAVTVEDKDLVNTANWRANYTILKGNENGNFKLIVTDAKTNEGVLVCVVKPLNVEEK 428
QY 523 -----TAAYSNIVFKAE--NPEPLVFGVKNASSFAKFTLIVTDVNEAPQFSQHVFOAKV 575
Db 429 QOMILQIGVNEAPFSREASPR-----SAMSTATVTVNVEDQDEGPECNPPITQVRM 480
QY 576 SEDVAIGTKVGNVTAKDPE---GLDISYSLRGDTRGWLKIDHVTGEIFSVAPLDREA--- 629
Db 481 KENAEVGTTSNGYKAYDPETRSRGIRYKLTDPDGTWMTIDENTGSIKVFSLDREAETI 540
QY 630 -GSPYRVQVVADEVGSSLSVSEFHLILMDVNDNPPRLAKDYTGTLFFCHPLSAPGSLIF 688
Db 541 KNGIYNITVLASDGGRTCTGT--LGIILODVNDNSPFIPK--KTVIICKPTMSSAEIV- 595
QY 689 EATDDQHLFRGPHFTFSLGSGS--LQNDWEYSKINGTHARLSRHTDTEERAYVVLIRI 746
Db 596 -AVDPDEPI-HGPPPDFSLESSTSEVQRMWRLKAINDTAARLSYQN-DPPPGSYVVPITV 652
QY 747 NDGGRPPLEGIVSLPVTFCSV-EGSCFRPAGHQGTPTVGMVAGILLTLLVIGIILAV 805
Db 653 RD--RLGMSSVTSILDVTLCDITENDCTHRVDPRIIGGGVGLGKWMAILAILLGIALLCFI 710
QY 806 VF 807
Db 711 LF 712

RESULT 15
IJHJHDB
N;Alternate names: desmosomal glycoprotein II
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 07-May-1999
C;Accession: B40390; S16465
R;Parker, A.E.; Wheeler, G.N.; Arnemann, J.; Pidsley, S.C.; Ataliotis, P.; Thomas, C.
J. Biol. Chem. 266, 10438-10445, 1991
A;Title: Desmosomal glycoproteins II and III. Cadherin-like junctional molecules gene
A;Reference number: A40390; MUID:91244819; PMID:2037591
A;Accession: B40390
A;Molecule type: mRNA
A;Residues: 1-901 <PAR>
A;Cross-references: GB:X56807
A;Note: it is uncertain whether Met-1 is the initiator or whether translation is init
R;Buxton, R.S.; Cowin, P.; Franke, W.W.; Garrod, D.R.; Green, K.J.; King, I.A.; Koch,
J. Cell Biol. 121, 481-483, 1993
A;Title: Nomenclature of the desmosomal cadherins.
A;Reference number: A40392; MUID:93252984; PMID:8486729
A;Contents: annotation; nomenclature
C;Genetics:
A;Gene: GDB:DSC3; DSC2; DSC1; DS
A;Cross-references: GDB:I26552; OMIM:600271
A;Map position: 18q12.1-18q12.1
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: alternative splicing; calcium binding; cell adhesion; duplication; glycop
F;1-28/Domain: signal sequence #status predicted <SIG>
F;136-901/Product: desmocollin 3a #status predicted <PRO>
F;136-695/Domain: extracellular #status predicted <EXT>
F;138-243/Domain: cadherin repeat homology <CRI>
F;246-355/Domain: cadherin repeat homology <CR2>
F;358-471/Domain: cadherin repeat homology <CR3>
F;474-577/Domain: cadherin repeat homology <CR4>
F;578-680/Domain: cadherin repeat homology <CR5>
F;696-718/Domain: transmembrane #status predicted <TM>
```

F:719-901/Domain: intracellular #status predicted <INT>  
F:166,392,546,629/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:864/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match		15.48;	Score 667.5;	DB 1;	Length 901;
Best Local Similarity		30.98;	Pred. No. 6.9e-34;		
Matches 186;		Conservative 104;	Mismatches 253;	Indels 59;	Gaps 21;
Qy	239	WKAKPKVEMVENSIDPHPIKIQVRWNPDAQYSL-----VDKEKLPRFPFSIDQE-	289		
Db	137	W-APIPCSMLENSLGPFLFLOQVQ-SDTAQNYIYISIRGPGVDQE--PRNLFYVERDT	192		
Qy	290	GDIYVTPQLDREEDKAYVYAVAKDEYCKPLSYPLEIHVKVKDINDNPPTCPSPVTVEV	349		
Db	193	GNLYCTRPVDRQEVESFIIAFAATPDGCTPELPLPLIIKIEDENDNYPFIETEYTFI	252		
Qy	350	QENERLGNISIGTLTAHDREDENTANSFLNYRIVEOTPKLPMDGLFLIQTYAGMLQAKQS	409		
Db	253	FENCVRGVTGVCATDKDEPDMHRLKYSIIGOVP--PSPTLFSMHPTTGVIITSSQ	310		
Qy	410	LKKQDTPOYNLTIEVSKD-----FKTLCFVQINVIDINDOIPIFEKSDYGNLTLAEDTN	464		
Db	311	LDRELIDKYQLKIKVODMDGOYFGLQTTSTCIIINDVDNDHLPFTFRSY--VTSVEENT	368		
Qy	465	IGSTILTQATDADEPFTGSSKILYHIHKDSEGRGLGVDTPHTNTGYVILKKPLDPE--	522		
Db	369	VDVEILRVTVEDKDLVNTANRANYTILKGNENGFKIVTDAKTNEGVLCVVKPLNYEEK	428		
Qy	523	-----TAAVSNIVFKAE-NPEPLVFGVKYNASSFAKETLIVTDVNEAPOFSQHVFOAKV	575		
Db	429	QOMILQIGVNEAPFSREASPR-----SAMSTATVTNVVEDODEGECNPPPIQTVRM	480		
Qy	576	SEDVAIGTKVGNVTAKDPE---GLDISYSLRGDTRGWLKIDHVTGEIPIFSVAPLDREA---	629		
Db	481	KENAEVGTTSNGYKAYDETRSSSGIRYKKLTDPTGWTTIDENTGSIKVFSLDREAETI	540		
Qy	630	-GSPYRVQVATEVGGSSLSVSEFHLILMDVNDNPPRLAKDYTGTLFCHPLSAPGSLIF	688		
Db	541	KNGIYNITVLASDQGRCTGT--LGIILQDVNDNSPFIPK--KTVIICKPTMSSAEIV-	595		
Qy	689	EATDDOHLFRGPHFTFSLGSS--LONDWEVSKINGTHARLSTRHTDTEERAYVVLIRI	746		
Db	596	-AVDDEPI-HGPPDFLESSTSEVQRMRLKALINDTAARLSYQN-DPPFGSYVVPITV	652		
Qy	747	NDGRRPPLLEGIVSLPVTFCSCV-EGSCFRPAGHOTGPTVGMVAGILLTLLVIGIILAV	805		
Db	653	RD--RLGMSSVTSLDVTLCDCTITENDCTHRVDPRIIGGGVQLGKWAILAILLGIALLF	710		
Qy	806	VF 807			
Db	711	LF 712			

Search completed: April 2, 2003, 15:54:34  
Job time : 35.3927 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 2, 2003, 15:48:04 ; Search time 17.0959 Seconds  
(without alignments)  
2018.515 Million cell updates/sec

Title: US-09-079-678-178

Perfect score: 4321

Sequence: 1 MILQAHLSLCLMLYLATG.....DRGKDNVESQAQSEVRPLRS 832

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB	ID	Description
1	4321	100.0	832	1	CADH_HUMAN	Q12864 homo sapien
2	3450	79.8	827	1	CADH_MOUSE	Q9r100 mus musculus
3	3414	79.0	827	1	CADH_RAT	P55281 rattus norv
4	954	22.1	829	1	CADG_RABIT	Q28634 oryctolagus
5	922.5	21.3	829	1	CADG_HUMAN	O75309 homo sapien
6	877.5	20.3	830	1	CADG_MOUSE	O88338 mus musculus
7	715	16.5	913	1	CAD4_CHICK	P24503 gallus gall
8	705	16.3	913	1	CAD4_MOUSE	P39038 mus musculus
9	699.5	16.2	906	1	CAD2_HUMAN	P19022 homo sapien
10	696	16.1	912	1	CAD2_CHICK	P10288 gallus gall
11	694.5	16.1	906	1	CAD2_MOUSE	P15116 mus musculus
12	688.5	15.9	877	1	CAD2_BOVIN	P19534 bos taurus
13	686.5	15.9	906	1	CAD2_RAT	O921v3 rattus norv
14	684.5	15.8	916	1	CAD4_HUMAN	P55283 homo sapien
15	674.5	15.6	902	1	DSC2_MOUSE	P55292 mus musculus
16	671.5	15.5	896	1	DSC3_HUMAN	Q14574 homo sapien
17	667.5	15.4	901	1	DSC2_HUMAN	Q02487 homo sapien
18	663.5	15.4	906	1	CAD0_XENLA	P33147 xenopus lae
19	661	15.3	896	1	DSC3_BOVIN	Q28060 bos taurus
20	657	15.2	713	1	CADD_HUMAN	P55290 homo sapien
21	653.5	15.1	863	1	DSC2_BOVIN	P33545 bos taurus
22	647.5	15.0	905	1	CADN_XENLA	P20310 xenopus lae
23	647	15.0	714	1	CADD_MOUSE	Q9wt5 mus musculus
24	642	14.9	895	1	DSC3_MOUSE	P5850 mus musculus
25	634	14.7	887	1	CADL_CHICK	P08641 gallus gall
26	630.5	14.6	712	1	CADD_CHICK	P33150 gallus gall
27	609	14.1	886	1	CADL_RAT	Q9r0t4 rattus norv
28	608	14.1	790	1	CADI_HUMAN	Q13634 homo sapien
29	605	14.0	4590	1	FATH_HUMAN	Q14517 homo sapien
30	604	14.0	884	1	CADL_MOUSE	P09803 mus musculus
31	601.5	13.9	3579	1	STAN_DROME	Q9v5n8 drosophila
32	597.5	13.8	732	1	CADL_CHICK	P33145 gallus gall
33	595.5	13.8	886	1	DSCI_MOUSE	P55849 mus musculus

RESULT 1

ID	CADH_HUMAN	STANDARD;	PRT;	832 AA.
AC	Q12864; Q15336;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Cadherin-17 precursor (Liver-intestine-cadherin) (LI-cadherin)			
DE	(Intestinal peptide-associated transporter HPT-1).			
GN	CDH17.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.			
RC	TISSUE=Colon adenocarcinoma;			
RX	MEDLINE=94204643; PubMed=8153632;			
RA	Dantzig A.H., Hoskins J., Tabas L.B., Bright S., Shepard R.L.,			
RA	Jenkins I.L., Duckworth D.C., Sportsman J.R., Mackensen D.,			
RA	Rostock P.R. Jr., Skatrud P.L.;			
RT	"Association of intestinal peptide transport with a protein related to			
RT	the cadherin superfamily.";			
RL	Science 264:430-433(1994).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Boettlinger A., Krefl B., Fieger C., Dlouhy B., Berndorff D.,			
RA	Goesner R., Tauber R.;			
RT	"Molecular cloning of human LI-cadherin: evidence for a novel type			
RT	of cadherin within the cadherin superfamily.";			
RL	Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.			
CC	-!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.			
CC	THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC			
CC	MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE			
CC	SORTING OF HETEROGENEOUS CELL TYPES. LI-CADHERIN MAY HAVE A ROLE			
CC	IN THE MORPHOLOGICAL ORGANIZATION OF LIVER AND INTESTINE.			
CC	INVOLVED IN INTESTINAL PEPTIDE TRANSPORT.			
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).			
CC	-!- TISSUE SPECIFICITY: EXPRESSED IN THE GASTROINTESTINAL TRACT AND			
CC	PANCREATIC DUCT. NOT DETECTED IN KIDNEY, LUNG, LIVER, BRAIN,			
CC	ADRENAL GLAND AND SKIN.			
CC	-!- SIMILARITY: CONTAINS 7 CADHERIN DOMAINS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; U07969; AAA19021.1; -;			
DR	EMBL; X83228; CAA58231.1; -;			
DR	HSSP; P15116; LNCJ.			
DR	Genew; HGNC:1756; CDH17.			
DR	MIM; 603017; -;			

DR InterPro; IPR002126; Cadherin.  
DR Pfam; PF00028; cadherin; 7.  
DR PRINTS; PR00205; CADHERIN.  
DR SMART; SM00112; CA; 5.  
DR PROSITE; PS00232; CADHERIN\_1; 3.  
DR PROSITE; PS0268; CADHERIN\_2; 6.  
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;  
KW Signal; Transport.  
FT CHAIN 1 22 POTENTIAL.  
FT CHAIN 23 832 CADHERIN-17.  
FT DOMAIN 23 787 EXTRACELLULAR (POTENTIAL).  
FT TRANSEM 788 808 POTENTIAL.  
FT DOMAIN 809 832 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 1 30 128 CADHERIN 1.  
FT DOMAIN 129 244 CADHERIN 2.  
FT DOMAIN 245 340 CADHERIN 3.  
FT DOMAIN 341 449 CADHERIN 4.  
FT DOMAIN 450 566 CADHERIN 5.  
FT DOMAIN 567 667 CADHERIN 6.  
FT DOMAIN 668 777 CADHERIN 7.  
FT CARBOHYD 149 184 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 419 419 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 456 456 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 546 546 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 587 587 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 722 722 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 115 115 E -> K (IN REF. 2).  
FT CONFLICT 446 446 I -> T (IN REF. 2).  
FT CONFLICT 734 734 D -> E (IN REF. 2).  
FT CONFLICT 739 739 A -> E (IN REF. 2).  
SQ SEQUENCE 832 AA; 92147 MW; C6DB6DE60E8A1277 CRC64;  
  
Query Match 100.0%; Score 4321; DB 1; Length 832;  
Best Local Similarity 100.0%; Pred. No. 7.1e-256;  
Matches 832; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 MILQAHLSCLLMYLYATGYGQEGKFSGLPKMTFSIYGEQPSQIIYQFKAAPPVTF 60  
Db 1 MILQAHLSCLLMYLYATGYGQEGKFSGLPKMTFSIYGEQPSQIIYQFKAAPPVTF 60  
  
Qy 61 ELTGTDNIFVIERGLLYNRALDRETRSHNLQVAALDANGIIVEGVPITIEVKDIN 120  
Db 61 ELTGTDNIFVIERGLLYNRALDRETRSHNLQVAALDANGIIVEGVPITIEVKDIN 120  
  
Qy 121 DNRPTFLQSKYEGSVRQNSRQKPLPYVYNATDLDPPATNGQLYQYVIQPLMNNVMYF 180  
Db 121 DNRPTFLQSKYEGSVRQNSRQKPLPYVYNATDLDPPATNGQLYQYVIQPLMNNVMYF 180  
  
Qy 181 QINNTGAISLTREGSQELNPAKNSYNLVTSVKDMGQSGNSFSDTTSVDIIYVTENIWK 240  
Db 181 QINNTGAISLTREGSQELNPAKNSYNLVTSVKDMGQSGNSFSDTTSVDIIYVTENIWK 240  
  
Qy 241 APKPYEMVENSTDRPIKITQVRNDPCAQYSLVDKEXLPRPFSIDQEGDIYVTOPLDR 300  
Db 241 APKPYEMVENSTDRPIKITQVRNDPCAQYSLVDKEXLPRPFSIDQEGDIYVTOPLDR 300  
  
Qy 301 EKDQAYFYAVAKDYGKPLSPYLIHVKKVDINDNPPTCPSPVTFVQNERLGNISG 360  
Db 301 EKDQAYFYAVAKDYGKPLSPYLIHVKKVDINDNPPTCPSPVTFVQNERLGNISG 360  
  
Qy 361 TLTAHREDEENTANSFLNRYVEQTPKLPMDGLFLIQTAYGMLQAKSLKKQDTPQYNL 420  
Db 361 TLTAHREDEENTANSFLNRYVEQTPKLPMDGLFLIQTAYGMLQAKSLKKQDTPQYNL 420  
  
Qy 421 TIEVSDKDFKTLCFQVQINVIDINOIPIFEKSDYGNLTAEEDTNGSTILLITQATDAEP 480  
Db 421 TIEVSDKDFKTLCFQVQINVIDINOIPIFEKSDYGNLTAEEDTNGSTILLITQATDAEP 480  
  
Qy 481 FTGSKILYHIKDGSEGRGLGVDPHPNTGVIKPKLPDEFETAAVSNIVFKAENPEPLV 540  
Db 481 FTGSKILYHIKDGSEGRGLGVDPHPNTGVIKPKLPDEFETAAVSNIVFKAENPEPLV 540

Qy 541 FGVKYNASSFAKFTLIIVTVNNEAPQFSQHVFAKVSEDVAIGTKVGNVTAKDPEGLDISY 600  
Db 541 FGVKYNASSFAKFTLIIVTVNNEAPQFSQHVFAKVSEDVAIGTKVGNVTAKDPEGLDISY 600  
  
Qy 601 SLRGDTRGWLKIDIVTGEIFSVAPLDREAGSPYRVQVYVATEVGGSSLSVSEFHLILMDV 660  
Db 601 SLRGDTRGWLKIDIVTGEIFSVAPLDREAGSPYRVQVYVATEVGGSSLSVSEFHLILMDV 660  
  
Qy 661 NDNPPRLAKDYTLGFFCHPLSAPGSLFEATDDQHLFRGPHFTFSLGSGSLQNDWEYSK 720  
Db 661 NDNPPRLAKDYTLGFFCHPLSAPGSLFEATDDQHLFRGPHFTFSLGSGSLQNDWEYSK 720  
  
Qy 721 INGTCHARLSTRHTDFEERAYVVLIRINDGRRPPLLEGIVSLVETFCSCVEGSCFRPAGHOT 780  
Db 721 INGTCHARLSTRHTDFEERAYVVLIRINDGRRPPLLEGIVSLVETFCSCVEGSCFRPAGHOT 780  
  
Qy 781 GIPTVGMVAGILLTLLVIGIILAVVFIRIKKDKGNVESQAQASEVKPLRS 832  
Db 781 GIPTVGMVAGILLTLLVIGIILAVVFIRIKKDKGNVESQAQASEVKPLRS 832

## RESULT 2

CADH\_MOUSE STANDARD; PRT; 827 AA.  
ID CADH\_MOUSE AC Q9R100;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Cadherin-17 precursor (Liver-intestine-cadherin) (LI-cadherin) (BILL-  
GN CDH17).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Swiss Webster; TISSUE=Intestine;  
RA Angres B., Kim L., Tauber R.;  
RT "Li-cadherin gene expression during intestinal development.";  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A., SEQUENCE OF 26-33; 52-58; 74-81; 117-123 AND  
RP 490-509, TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.  
RC STRAIN=BALB/c; TISSUE=Fetal liver;  
RX MEDLINE=20469471; PubMed=10906147;  
RA Ohnishi K., Shimizu T., Karasuyama H., Melchers F.;  
RT "The identification of a nonclassical cadherin expressed during B cell  
development and its interaction with surrogate light chain.";  
RL J. Biol. Chem. 275:31134-31144(2000).  
CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.  
THEY PREPARENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC  
MANNER IN CONNECTING CELLS. CADHERINS MAY THUS CONTRIBUTE TO THE  
SORTING OF HETEROGENEOUS CELL TYPES. LI-CADHERIN MAY HAVE A ROLE  
IN THE MORPHOLOGICAL ORGANIZATION OF LIVER AND INTESTINE.  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).  
CC -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IS FOUND IN INTESTINE WITH  
LOWER EXPRESSION IN SPLEEN, BONE MARROW, LUNG AND TESTIS. NO  
EXPRESSION DETECTED IN LIVER, KIDNEY, HEART, BRAIN OR SKELETAL  
MUSCLE. EXPRESSED IN PRECURSOR B-CELLS AND MYELOID CELLS.  
CC -!- DEVELOPMENTAL STAGE: EXPRESSION INCREASES IN PRO- AND PRE-B-I  
CELLS, DECREASES IN LARGE AND SMALL PRE-B-II CELLS, AND INCREASES  
AGAIN IN IMMATURE AND MATURE B-CELLS.  
CC -!- SIMILARITY: CONTAINS 7 CADHERIN DOMAINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AF177669; AAD51125.1; -.  
 DR EMBL: D87912; BAB03264.1; -.  
 DR HSP: P15116; INCI.  
 DR MGD: MGI:1095414; Cdh17.  
 DR InterPro: IPR002126; Cadherin.  
 DR Pfam: PF00028; cadherin; 7.  
 DR PRINTS: PR00205; CADHERIN.  
 DR SMART: SM00112; CA; 5.  
 DR PROSITE: PS00232; CADHERIN\_1; 3.  
 DR PROSITE: PS0268; CADHERIN\_2; 5.  
 DR Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;  
 KW Signal; Transport.  
 FT SIGNAL 1 25  
 FT CHAIN 26 827  
 FT DOMAIN 26 786  
 FT TRANSMEM 787 807  
 FT POTENTIAL.  
 FT DOMAIN 808 827  
 FT CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 1. 129 127  
 FT CADHERIN 1.  
 FT DOMAIN 128 243  
 FT CADHERIN 2.  
 FT DOMAIN 244 339  
 FT CADHERIN 3.  
 FT DOMAIN 340 448  
 FT CADHERIN 4.  
 FT DOMAIN 449 565  
 FT CADHERIN 5.  
 FT DOMAIN 566 666  
 FT CADHERIN 6.  
 FT DOMAIN 667 776  
 FT CADHERIN 7.  
 FT CARBOHYD 148 148  
 FT N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 249 249  
 FT N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 418 418  
 FT N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 545 545  
 FT N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 573 573  
 FT N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 586 586  
 FT N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 721 721  
 FT N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 827 AA; 91645 MW; CDCEEA5A76E2B58 CRC64;  
 Query Match 79.88; Score 3450; DB 1; Length 827;  
 Best Local Similarity 79.28; Pred. No. 8.5e-203;  
 Matches 658; Conservative 75; Mismatches 94; Indels 4; Gaps 1;  
 QY 2 ILQAHLSLCLMLLYLATYGEQKSGPLKPMPTFSIEGQEPQIIIFQKPNPAPTVE 61  
 DB 1 MVSQALHFLCLLTLYLCGYGEEKSGPLKPMPTFSIEGQEPQIIIFQKPNPAPTVE 60  
 QY 62 LTGETDNFVIERGLLYNRLDRETRSHNLQVAALDANGIIEGVPVITIEVKDND 121  
 DB 61 LTGETDGIETKEDGLLYHTRALDRETRAVHHLQALDLSHGAVDGPVPTIEVKDND 120  
 QY 122 NRPTFLOSKYEGSVQRNCRKPLLYNATDLDLPATPNGOLYQYIVQLPNNVMYFQ 181  
 DB 121 NRPTFLOSKYEGSVQRNCRKPLLYNATDLDLPATPNGOLYQYIVQLPNNVMYFQ 180  
 QY 182 INKKTGAISLTREGSQELNPAKPNLYNLSVDMGQSGNSFSDTTSVDIIVTENTWKA 241  
 DB 181 IDSKTGAISLTREGSQELNPAKPNLYNLSVDMGQSGNSFSDTTSVDIIVTENTWKA 240  
 QY 242 PKPVMYENSTDPHPKITQVWRNDGAQYSLVDKELPRPFPFSDIGEDYIVVTPDLRE 301  
 DB 241 PEPVEIRENSTDPHPKITQVWRNDGAQYSLVDKELPRPFPFSDIGEDYIVVTPDLRE 300  
 QY 302 EKDAYVYAVAKDEYKPLSPLVHVKVDINDNPTCPSPVTVFVEQNERLGNSTGT 361  
 DB 301 EKNHVFATKDEYKPLSPLVHVKVDINDNPTCPSPVTVFVEQNERLGNSTGT 360  
 QY 362 LTAHREDEANTANFLAYRIVEQTPKPLMDGLFLTIQYAGMLQALSKKQDTPQYNLT 421  
 DB 361 FEAHMDEANNISILYKLVQDPKVPDGLFLIGEYKQVLSKSKKQDTPQYNLT 420  
 QY 422 IEVSDKFKTLCTFQVINDINDIPIFESKDYGNLNLAEADTNTGSTITLIQATDADEPF 481  
 DB 421 IEVSDVDFKTLCTYQVNVINDINDIPIFESKDYGNLNLAEADTNTGSTITLIQATDADEPF 480  
 QY 482 TGSSEKILYHIKGSBGRGLVDTPHNTGVYIHKPLDFTAAVSNIVFAENPEPLVF 541  
 DB 481 TGSSEKILYHIKGSBGRGLVDTPHNTGVYIHKPLDFTAAVSNIVFAENPEPLVF 540

QY 542 GVKYNASSFAKFTLLIVDVNEAPQFSQHVFAQKYSVEDVAIGTKVGNVYAKDPEGLDTSYS 601  
 DB 541 GIEYNASSFASFEIIVDVNEVPVQRIQFQANVSEDAAGVSGVGNVYAKDPEGLDTSYS 600  
 QY 602 LRGTGRLWKIDHVTGEIFSVAPLDREAGSPYRVQVYVATEVGGSSLSVSEFHLIMDVN 661  
 DB 601 LKGMRLWKIDHVTGEIFSVAPLDREAGSPYRVQVYVATEVGGSSLSVSEFHLIMDVN 660  
 QY 662 DNPRLAKDYTGTFCHPLSAPGSLIFEATDDOHLFRGHFTFSLGSLQNDWEYSKI 721  
 DB 661 DNPRLAKDYTGTFCHPLSAPGSLIFEATDDOHLFRGHFTFSLGSLQNDWEYSKI 720  
 QY 722 NGTHARLSTRHTRFEERAYVYLIRINDGRPPLEGIVSLPVTFCSEVGGSCFRPAGHOTG 781  
 DB 721 NGTHARLSTRHTRFEERAYVYLIRINDGRPPLEGIVSLPVTFCSEVGGSCFRPAGHOTG 780  
 QY 782 IPTVGMAGVILLTLLVIGIILAVVFIIRIKKDKGNDVESAQAQSEVAPLRS 832  
 DB 781 IPTVGMAGVILLTLLVIGIILAVVFIIRIKKDKGNDVESAQAQSEVAPLRS 832  
 RESULT 3  
 ID CADH\_RAT STANDARD; PRT; 827 AA.  
 AC P55281;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Cadherin-17 precursor (Liver-intestine-cadherin) (LI-cadherin).  
 GN CDH17.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Liver;  
 RX MEDLINE=94266966; PubMed=8207063;  
 RA Berndorff D., Gessner R., Kreft B., Schnoy N., Lajous-Petter A.-M.,  
 RA Loch N., Reutter W., Hortsch M., Tauber R.;  
 RT "Liver-intestine cadherin: molecular cloning and characterization of  
 RT a novel Ca(2+)-dependent cell adhesion molecule expressed in liver  
 RT and intestine.";  
 RL J. Cell Biol. 125:1353-1369(1994).  
 CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.  
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC  
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE  
 CC SORTING OF HETEROGENEOUS CELL TYPES. LI-CADHERIN MAY HAVE A ROLE  
 CC IN THE MORPHOLOGICAL ORGANIZATION OF LIVER AND INTESTINE.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- TISSUE SPECIFICITY: LIVER AND INTESTINE.  
 CC -!- SIMILARITY: CONTAINS 7 CADHERIN DOMAINS.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X78997; CAA55631.1; -.  
 DR HSP: P15116; INCI.  
 DR InterPro: IPR002126; Cadherin.  
 DR Pfam: PF00028; cadherin; 7.  
 DR PRINTS: PR00205; CADHERIN.  
 DR SMART: SM00112; CA; 6.  
 DR PROSITE: PS00232; CADHERIN\_1; 3.  
 DR PROSITE: PS0268; CADHERIN\_2; 6.  
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;  
 KW Signal.  
 FT SIGNAL 1 21  
 FT POTENTIAL.

```
FT CHAIN 22 827 CADHERIN-17.
FT DOMAIN 22 786 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 787 807 POTENTIAL.
FT DOMAIN 808 827 CYTOPLASMIC (POTENTIAL).
FT CHAIN 29 127 CADHERIN 1.
FT DOMAIN 128 243 CADHERIN 2.
FT DOMAIN 244 339 CADHERIN 3.
FT DOMAIN 340 448 CADHERIN 4.
FT DOMAIN 449 565 CADHERIN 5.
FT DOMAIN 566 666 CADHERIN 6.
FT DOMAIN 667 776 CADHERIN 7.
FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 183 183 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 249 249 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 545 545 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 721 721 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 827 AA; 91862 MW; E1A0B03646562C01 CRC64;

Query Match 79.0%; Score 3414; DB 1; Length 827;
Best Local Similarity 78.8%; Pred. No. 1.3e-200;
Matches 655; Conservative 70; Mismatches 102; Indels 4; Gaps 1;

Qy 2 ILQAHLSLCLMLLYLATGYGQEGKFSGLPKPMTFISIYEGQEPQIIPQFKANPPAVTFE 61
Db 1 MVSQAHLFLCLTLYLATGYGQEGKFSGLPKPMTFISIFEGQEPQIIPQFKANPPAVTFE 60

Qy 62 LTGETDNFVIEREGLYNRLDRETRSTHNLQVAALDANGIIVGVPVITIEVKDND 121
Db 61 LTGETDGLFKTEKDGLLYHTRVLDRETRAVHQLAALDSQAIYVGVPIIIEVKDND 120

Qy 122 NRPTFLOSKYEGSVQRNRPKPELYVNATDLDPPATNGQLYQIVYIQLPMINNMVYFQ 181
Db 121 NRPTFLOSKYEGSVQRNRPKPELYVNATDLDPPATNGQLYQIVYIQLPMINNMVYFQ 180

Qy 182 INNKTAISLTPREGSOELNPAKNPSYLNIVSKDMGGQSENFSFDTTSVDIIVTENIWA 241
Db 181 INNKTAISLTPREGSOELNPAKNPSYLNIVSKDMGGQSENFSFDTTSVDIIVTENIWA 240

Qy 242 PKPVMENSTDPHPKIKTQVWMDPGAGYSLVADKEKLPREFPSIDQEGDIYVTPQLDRE 301
Db 241 PEPVEIRENLDPHPKIKTQVWMDPGAGYSLVADKEKLPREFPSIDQEGDIYVTPQLDRE 300

Qy 302 EKDAYVFAVAKDEYKGLPSTPLEIHKVVKOINDNPPCPSPVTVFVEQNERLGNSTGT 361
Db 301 EKDSHVFFATAKDENGKPLAYPLEIRVVRVIDINDNPPCLSQVTVFVEQNEVLGSSIGI 360

Qy 362 LTAHDDRENTANSPLNRYIVEQTPKPMQDGLFLQTYAGMLQAKOSLKKODTPOYNLT 421
Db 361 FAHDDRENTANSPLNRYIVEQTPKPMQDGLFLQTYAGMLQAKOSLKKODTPOYNLT 420

Qy 422 IEVSKDKFTLCFQVQINVIDINDQIPIPEKSDYGNLTLAEDNTIGSTLTITQATDADPEF 481
Db 421 VEVSDIDFKTLCSLQVNVINDINDQIPIPEKSDYGNLTLAEDNTIGSTLTITQATDADPEF 480

Qy 482 TGSSKILYHIKGOSEGLRGVDTDPHTNTGYVIKKPLDFETAASVNFVKAENPEPLVF 541
Db 481 TGSSKILYHIKGOSEGLRGVDTDPHTNTGYVIKKPLDFETAASVNFVKAENPEPLVF 540

Qy 542 GVKYNASSFAKFTLIVTDVNEAPQSFVHFOAKVSEDVAIGTKVGNVTAQDPEGLDISY 601
Db 541 GVKYNASSFAKFTLIVTDVNEAPQSFVHFOAKVSEDVAIGTKVGNVTAQDPEGLDISY 600

Qy 602 LRGDPRGWLKIDHVTGIFSVAPLDREAGSPYRVQVAVTEVGGSSLSVSEFHLTMDVN 661
Db 601 LRGDPRGWLKIDHVTGIFSVAPLDREAGSPYRVQVAVTEVGGSSLSVSEFHLTMDVN 660

Qy 662 DNPPLAKDYTGFLFCHPLSPAGSLIFEATDDQHLFRPHFTFSLGSGSLQNDWEVSKI 721
Db 661 DNPPLAKDYTGFLFCHPLSPAGSLIFEATDDQHLFRPHFTFSLGSGSLQNDWEVSKI 720

Qy 722 NGTHARLSTRHDTFERAYVYVILIRINDGRPPLEGIVSLPVTFCVSGVSCFRPAGHTG 781
```

```
Db 721 NGTHARLSTRHDTFERAYVYVILIRINDGRPPLEGIVSLPVTFCVSGVSCFRPAGHTG 780
Qy 782 IPTVGMAGVIGLLTLLVIGIILAVFVIRIKDKGKNVESAQASEVVKPLRS 832
Db 781 IPTVGMAGVIGLLTLLVIGIILAVFVIRIKDKGKNVESAQASEVVKPLRS 827

RESULT 4
CADG_RABIT STANDARD; PRT; 829 AA.
ID CADG_RABIT
AC Q28634;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE cadherin-16 precursor (Kidney-specific cadherin) (Ksp-cadherin).
GN CDH16.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A., AND OF SEQUENCE 570-586.
RC STRAIN=New Zealand white;
EX MEDLINE=95340560; PubMed=7615566;
RA Thomson R.B., Igarashi P., Biemesderfer D., Kim R., Abu-Alfa A.,
RA Soleimani M., Aronson P.S.;
RT "Isolation and cDNA cloning of ksp-cadherin, a novel kidney-specific
member of the cadherin multigene family.";
RL J. Biol. Chem. 270:17594-17601(1995).
CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: KIDNEY-SPECIFIC. LIMITED TO THE BASOLATERAL
CC MEMBRANES OF RENAL TUBULAR EPITHELIAL CELLS.
CC -!- SIMILARITY: CONTAINS 6 CADHERIN DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U28945; AAC48472.1; -
CC HSP; P09803; 1SUH.
CC InterPro: IPR002126; Cadherin.
CC Pfam; PF00028; cadherin; 6.
CC SMART; SM00112; CA; 6.
CC PROSITE; PS00232; CADHERIN_1; 2.
CC PROSITE; PS0268; CADHERIN_2; 6.
CC Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
CC Signal.
KW SIGNAL 1 18 POTENTIAL.
KW CHAIN 19 829 CADHERIN-16.
KW DOMAIN 19 786 EXTRACELLULAR (POTENTIAL).
KW TRANSMEM 787 807 POTENTIAL.
KW DOMAIN 808 829 CYTOPLASMIC (POTENTIAL).
KW DOMAIN 25 126 CADHERIN 1.
KW DOMAIN 131 235 CADHERIN 2.
KW DOMAIN 242 336 CADHERIN 3.
KW DOMAIN 341 449 CADHERIN 4.
KW DOMAIN 455 564 CADHERIN 5.
KW DOMAIN 569 665 CADHERIN 6.
KW DOMAIN 666 786 ECTODOMAIN G.
KW CARBOHYD 517 517 N-LINKED (GLCNAC. . .) (POTENTIAL).
KW CARBOHYD 602 602 N-LINKED (GLCNAC. . .) (POTENTIAL).
KW CARBOHYD 709 709 N-LINKED (GLCNAC. . .) (POTENTIAL).
KW CARBOHYD 722 722 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 829 AA; 88827 MW; D2FD10E6C47A43B9 CRC64;
```

```
Query Match      22.1%; Score 954; DB 1; Length 829;
Best Local Similarity 31.6%; Pred. No. 1.1e-50;
Matches 251; Conservative 142; Mismatches 354; Indels 48; Gaps 22;

Qy 62 LTGET----DNIFVIERE-GLLYNRRALDRSTRTHNQVAALDANGIIVGPGVPITIEV 116
Db 57 LSGDSGVAAGPFFVEAGSEGLLVTRALDREAEQYQVTLAEADGHLVWGQSVTVHV 116
Qy 117 KDINDNRPTEFQSKVEGVSVRNSRPGKPLVYNATDLDPATNGQLYQVIVQLPMINN 176
Db 117 KDENDQVPHFQSAIYRRLSRTGTPDPLFLEASDRDEPGTANSLRHLISQPAQPS 176
Qy 177 VMYFQINNKTGAISLTREGSQELNPAKNPSYLVISVKMGQSGNSFSDTTSVDIIVTE 236
Db 177 PDMQLEPRICALALSPKGSLSLHALERTYQLLVQVKMDGQA-SGHQATATVEVSIIE 235
Qy 237 NIWKAPRPMVENSTDPHPKIKITQVRNDPGCAQSLVDKELPRFPFSIDQEGDIYVTO 296
Db 236 STWTSLEPIHLAENLVLYPHHMAQVHWSGGDVHYHL---ESHPPGPPEVNAEGNLYVTR 292
Qy 297 PLDREEDKAVYFAVAKDEYCKPLSYPLEIHHVKVDINDNPPTCP---SPVTVEVQENE 353
Db 293 ELDRAGQGYVQVQAQNSRGEDYAEPLHLHVVTVDNDHAPVCPPRGPPVSVPELSP- 351
Qy 354 RLNSIGTTLTAHREDEANTSNFLNRYIVEQTPKL-PMDGLFLIQTIVAGMLQAKQSLKK 412
Db 352 --GTAVTTLSEADADACPSNHHVYRLSPPEQEGPQEGAFQDPTSGSVSLGAAPLEA 409
Qy 413 QDTQYNLTIEV-----SKDKFTKTCFQVINDINDQIPIFEKSDYGNLTAEADTNI 465
Db 410 GQ-----NMLLQVLAADLAGAAGLSSTCEVAVTVTDVNDHAPFTSSQVGPVSLPEDTEP 465
Qy 466 GSTILTQATDAD-EPFTGSSKIL-YHIIKGDSEGRGLGVDPDHTNCTGYIIKKPLDPEET 523
Db 466 GTLVATLTATDADLEP---APRLMDFTIEAGDGGTGLDWEPSGHVQLYLLKNLSYEA 522
Qy 524 AAVSNIVFAENPEPLVFGVKYNASSFAKFTLIIVTDVNEAPQFSQHVQFQAKVSEDVAIGT 583
Db 523 APSHTVVVVVNRVYETV-GPGPGGATATVTVLVEKVMPPRLEQKSVEADIPVNPAPGS 581
Qy 584 KGVNTAKDPGLDISVSLRQDTGWLKIDHVTGEISVAPLD-REAGSPRVQVATEV 642
Db 582 FLLTIQPAEPNGWALRFSLVNDSEGVFCIQKVSSEVHTARPLQGARPGDSYTVLVEAQDA 641
Qy 643 GGSLSVSEPHLLIMDVNDNPPRLAKDYTGLEFCHPLSAPGSLIFEATDD-DQHLRGP 701
Db 642 DAPRLSTSAALVIHFLRAPAPALPLAPMPSRHLCPTRODHGVLIIPAPSEDPDMATGHP 701
Qy 702 HFTFSLG-SGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSL 760
Db 702 -YSPALGNPTVQRDRLQLNDLSHAFTLALHWVEPREHIVPVVVSQDAR-----VWQL 755
Qy 761 P--VTFCS-VEGSCFRPAGHOTGIPVGMVAGLILLTLLVIGIILAVFTRIKDKGKD 817
Db 756 PVRVVVCRNTEGECMRKVRGMKGMPTKLSAVGILVGTLLAIGFLLIFLTHALARKKD 815
Qy 818 NVESAQAASEVKPLRS 832
Db 816 --LDAPADNV-PLKA 827

RESULT 5
CADG_HUMAN STANDARD; PRT; 829 AA.
AC O75309;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cadherin-16 precursor (Kidney-specific cadherin) (Ksp-cadherin).
GN CDH16.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

Qy 297 PLDREKDAVYFAVAKEDEYKPLSYPLEIHVKVKVDINDNPPCTCPSPVTVFVEQNERLG 356  
 Db 293 ELDRQAQVLLQVRAQNSHGEDIAAPLHLVLMVDENDNVPICPPRDPVTVPISPELSPG 352  
 Qy 357 NSIGTLTAHDEDEANTANFLNYRIVBOTKPLMDG-LFLIQTAYGMLQAKOSLKK-QD 414  
 Db 353 TEVTRLSAEDADPGSPNSHVYVYLLSPEDGVEGRAFDVPTSGSVTLGLVPLRAGQ 412  
 Qy 415 TPQNLNIEVSKD--FKTLCFVQINVIDINDQIPFEKSDYGNLTLAEDNIGSTIITI 472  
 Db 413 ILLVLAMLAGAGGFSSTCEVEAVTDINDHAPEFITSQIGFISLPEDVEPGTLVAML 472  
 Qy 473 QATDAD-EPPTGSSKIL-YHIIKDSGRGLGVDTPHTNTGYVIK--KPLDFTAAVSN 528  
 Db 473 TAIDADLEP---AFRLMDFAIERGDTGTFGLDWEF--DSGHVRLCKNLSEYEAAPSHE 527  
 Qy 529 IVFKAENPEPLVFCVKYNASSFAKFTLVTDVNEAPOSQHVFOAKYSEDVAIGTKYGNV 588  
 Db 528 VVVVVQVAKLV-GPGGPGGATATVTLVERMPPPKLDQESYASVPISAPAGSFLITI 586  
 Qy 589 TAKDPEGLDISLRGDTGRGLKIDHTVTELFVAPLD-REAGSPYRVQVVAEVEGSSL 647  
 Db 587 QPSDPSIRTLFSLVNDSEGWLCTEKSSEVHTAQSLQAGPGDTYVVLVEAQDTDEPL 646  
 Qy 648 SSVSEFHLIMDVNDNPPRLAKDYTGTLFFCHPLSAPGLIP-EATDDQHLFRGPHPTFS 706  
 Db 647 SASAPLVTHFLKAPPAPALTAPVPSQVLCYPRQDHGLIVSGPSKDPDLASGHGP-YSFT 705  
 Qy 707 LG-SGSLQNDKEVSKINGTHARLSTRHTDDEERAVVVLIRINDGRPLEGIVSLPVTF 765  
 Db 706 LGPNTVQVORQRLQTLNGSHAYLTLALHWVPEHRIIPVVVSHNAQ---MQLLVLRVIVC 762  
 Qy 766 SC-VEGSCFRPAGHTGPTVGMAGILLTLLVTGILAVVFIIRIKDKGKNVESQA 824  
 Db 763 RCNVEGQCMRKVGKMGKPTKLSAGVILVGLTVAIGFLLIFLTHWTMSRKKDPQPADS 822  
 Qy 825 SEVK 828  
 Db 823 VPLK 826

RESULT 6  
 CADG\_MOUSE STANDARD; PRT; 830 AA.  
 AC 088338; 09JLZ5;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE cadherin-16 precursor (kidney-specific cadherin) (Ksp-cadherin).  
 GN CDH16.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 ON NCBI\_TaxID:10090;  
 RX [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=BALB/C;  
 RX MEDLINE=98389630; PubMed=9721215;  
 RA Thomson R.B., Ward D.C., Quaggin S.E., Igarashi P., Muckler Z.E.,  
 RA Anonson P.S.;  
 RT "cDNA cloning and chromosomal localization of the human and mouse  
 RT isoforms of Ksp-cadherin."  
 RL Genomics 51:445-451(1998).  
 RP [2]  
 RP SEQUENCE OF 1-25 FROM N.A.  
 RC STRAIN=129/Sv;  
 RA Whyte D.A., Thomson R.B., Nix S.L., Zanjan R., Li C., Karp S.L.,  
 RA Anonson P.S., Igarashi P.;  
 RT "Ksp-cadherin gene promoter. I. Sequence analysis and renal  
 RT epithelial-cell-specific activity."  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.

CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC  
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE  
 CC SORTING OF HETEROGENEOUS CELL TYPES.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Probable).  
 CC -1- TISSUE SPECIFICITY: KIDNEY-SPECIFIC.  
 CC -1- SIMILARITY: CONTAINS 6 CADHERIN DOMAINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@sib-sib.ch).  
 CC  
 CC EMBL; AF016271; AAC34254.1; -;  
 DR EMBL; AF118228; AAF28836.1; -;  
 DR HSP; P09803; 1SUH.  
 DR GSD; MGI:106671; Cdh16.  
 DR InterPro: IPR002126; Cadherin.  
 DR Pfam; PF00028; cadherin; 6.  
 DR SMART; SM00112; CA; 6.  
 DR PROSITE; PS00232; CADHERIN\_1; 1.  
 DR PROSITE; PS00268; CADHERIN\_2; 5.  
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;  
 KW Signal.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 830 CADHERIN-16.  
 FT DOMAIN 22 788 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 789 809 POTENTIAL.  
 FT DOMAIN 810 830 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 27 128 CADHERIN 1.  
 FT DOMAIN 133 237 CADHERIN 2.  
 FT DOMAIN 244 338 CADHERIN 3.  
 FT DOMAIN 343 451 CADHERIN 4.  
 FT DOMAIN 457 566 CADHERIN 5.  
 FT DOMAIN 571 667 CADHERIN 6.  
 FT DOMAIN 668 788 ECTODOMAIN G.  
 FT CARBOHYD 519 519 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 604 604 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 724 724 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 830 AA; 89859 MW; 5EB1F06B5B7C2F25 CRC64;  
 Query Match 20.3%; Score 877.5; DB 1; Length 830;  
 Best Local Similarity 30.5%; Pred No. 5.1e-46;  
 Matches 245; Conservative 141; Mismatches 333; Indels 63; Gaps 22;  
 Qy 62 LTGETD----NIFVIERE-GLLYYNRALDRETRSTHNLQVAALDANGIIVEGVPITIEV 116  
 Db 59 LSGDSNTAQNTFAVDTSDGFLVATRTLDREKAQYQLQVLTSESDGRLWGPQLVTVHV 118  
 Qy 117 KDINDNRTFFLOSKYEGSVRONSPGKFLYVYNATDLDPATPNQLYQIVQLPMINN 176  
 Db 119 KDENDQVQFQSAIYRAQLSQCTRGVPFLFLEASDGPAGCTANSDLRHLISQSPQPL 178  
 Qy 177 VMYFQINKTKAISLTREGSQELNPAKNPSYNLVLSVKDMGSGSENSESDTSDVDIIVE 236  
 Db 179 PDMFQDLPHLGALALSPSGSTSLDHALDEETVQLLVQVMDMGDQPSGHQAIAT-VEISIVE 237  
 Qy 237 NIWKAPKVEVVENSTDPHPKIKITQVRWNPQCAQYSLVDKPKLPFPFSDIQEGDIYVQ 296  
 Db 238 NSWAPLEPVHLAENLKVVYPHSIAQVHSGGDVHYQL---ESQPPGPDVDTEGLHVTM 294  
 Qy 297 PLDREKDAYVYFAVAKDEYKPLSYPLEIHVKVKVDINDNPPCTCPSPVTVFVEQNERLG 356  
 Db 295 ELDRQAQVLLQVRAQNSHGEDIAAPLHLVLMVDENDNVPICPPRDPVTVPISPELSPG 354  
 Qy 357 NSIGTLTAHDEDEANTANFLNYRIVBOTKPLMDG-LFLIQTAYGMLQAKOSLKK-QD 409  
 Db 353 TEVTRLSAEDADPGSPNSHVYVYLLSPEDGVEGRAFNKAFELDPTSGSVTLGTAPLHAGQS 414  
 Qy 410 LKKQDTPQYNLTIEV--SDKDFKTLFCFVQINVIDINDQIPFEKSDYGNLTLAEDNIGTS 467







RESULT 9

CAD2\_HUMAN STANDARD; PRT; 906 AA.  
 AC P19022; O14923;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Neural-cadherin precursor (N-cadherin) (Cadherin-2).  
 GN CDH2 OR CDHN OR NCAD.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91016946; PubMed=2216790;  
 RA Reid R.A., Hemperly J.J.;  
 RT "Human N-cadherin: nucleotide and deduced amino acid sequence.";  
 RL Nucleic Acids Res. 18:5896-5896(1990).  
 RN [2]  
 RP REVISIONS TO 341; 699 AND 705.  
 RA Reid R.A.;  
 RT Submitted (NOV-1990) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92363956; PubMed=1500442;  
 RA Salomon D., Ayalon O., Patel-King R., Hynes R.O., Geiger B.;  
 RT "Extrajunctional distribution of N-cadherin in cultured human  
 endothelial cells";  
 RL J. Cell Sci. 102:7-17(1993).  
 RN [4]  
 RP SEQUENCE OF 160-906 FROM N.A.  
 RX MEDLINE=90347462; PubMed=2384753;  
 RA Walsh F.S., Barton C.H., Putt W., Moore S.E., Kelsell D.,  
 Spurr N., Goodfellow P.N.;  
 RT "N-cadherin gene maps to human chromosome 18 and is not linked to the  
 E-cadherin gene.";  
 RL J. Neurochem. 55:805-812(1990).  
 RN [5]  
 RP SEQUENCE OF 1-20 FROM N.A.  
 RX MEDLINE=95048366; PubMed=7959764;  
 RA Wallis J.A., Fox M., Walsh F.S.;  
 RT "Structure of the human N-cadherin gene: YAC analysis and fine  
 chromosomal mapping to 18q11.2";  
 RL Genomics 22:172-179(1994).  
 CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.  
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC  
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE  
 CC SORTING OF HETEROGENEOUS CELL TYPES. N-CADHERIN MAY BE INVOLVED IN  
 CC NEURONAL RECOGNITION MECHANISM.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X57548; CAA04773.1; -  
 CC EMBL; X54315; CAA38213.1; -  
 CC EMBL; S42303; AAB22854.1; -  
 CC EMBL; M34064; AAA03236.1; -  
 CC EMBL; Z27420; CAA81799.1; -  
 CC PIR; A38870; IJHUCN.  
 CC HSSP; P15116; INCUJ.  
 CC Genew; HGNC:1759; CDH2.  
 CC MIM; 114020; -  
 CC InterPro; IPR002126; Cadherin.  
 CC InterPro; IPR002333; Cadherin\_C\_term.

DR Pfam; PF00028; cadherin; 5.  
 DR Pfam; PF01049; Cadherin\_C\_term; 1.  
 DR PRINTS; PRO0205; CADHERIN.  
 DR SMART; SM00112; CA; 5.  
 DR PROSITE; PS00232; CADHERIN\_1; 3.  
 DR PROSITE; PS02068; CADHERIN\_2; 5.  
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;  
 KW Signal.  
 FT SIGNAL 1 23 POTENTIAL.  
 FT PROPEP 24 159  
 FT CHAIN 160 906 NEURAL-CADHERIN.  
 FT DOMAIN 160 724 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 725 746 POTENTIAL.  
 FT DOMAIN 747 906 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 160 267 CADHERIN 1.  
 FT DOMAIN 268 382 CADHERIN 2.  
 FT DOMAIN 383 497 CADHERIN 3.  
 FT DOMAIN 498 603 CADHERIN 4.  
 FT DOMAIN 604 714 CADHERIN 5.  
 FT DOMAIN 714 863 SER-RICH.  
 FT CARBOHYD 190 190 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 273 273 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 325 325 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 402 402 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 572 572 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 651 651 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 692 692 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CONFLICT 12 12 MISSING (IN REF. 3).  
 FT CONFLICT 16 16 L -> A (IN REF. 3 AND 5).  
 FT CONFLICT 196 196 S -> T (IN REF. 1).  
 FT CONFLICT 212 212 I -> L (IN REF. 4).  
 FT CONFLICT 357 357 N -> I (IN REF. 1).  
 FT CONFLICT 867 867 A -> L (IN REF. 3).  
 SQ SEQUENCE 906 AA; 72DDC7B8B57C7AFC CRC64;  
 Query Match 16.2%; Score 699.5; DB 1; Length 906;  
 Best Local Similarity 27.4%; Pred. No. 4.le-35;  
 Matches 210; Conservative 122; Mismatches 331; Indels 103; Gaps 21;  
 QY 117 KDINDNRPFTLQSGVSRGPKFLVYNATDLDPATPNGLQYQVIVQIPLMNN 176  
 DB 47 KDVEGQP-LLNVKF-----SNCNGKRVQYSESEPADFKVDEDMGVY--AVRSEPLSE 98  
 QY 177 VMYFQINNKTGAISLTREGSQELNPAKNPSYLVISVKDMGQSGNSFSDTTSVDIIV-- 234  
 DB 99 HAKELIV-----AQDKETQEKQVAVKLSLAPT--LTEESVKESAEEVIFP 144  
 QY 235 -----TENIKAPKPVEMVENSTDPHPKIKITQVRWN-----DPGAQ 270  
 DB 145 RQSKHSGHLQQRQDWVIP-PINLPENSRGPFQELVRIISRDRDKNLSLRYSVTGPGA- 202  
 QY 271 YSLVDKEKLPRFPSIDQ-EGDIYVTOPLDREKDAYVYAVAKDEYKPLSYPLEIHVK 329  
 DB 203 -----DQPPGTGIINPISQGLSVTKPLDREQIARFHLRAHVDINGNQVENPIDVIN 256  
 QY 330 VKDINDNPPCTSPVPTVEVQENRGLNSIGTLTAHDREENTANSFLNRYRIVEQTPKLP 389  
 DB 257 VIDMNDNRPEFLHQNCTVPEGSKPGTYVMTVTAIDADDPNALNGMLRYRIVSQAPSTP 316  
 QY 390 MDGLFIQTYAGMLQLAKQSLKQDTPQYNTLIEYKDK-----FKTLCFVQINVIDIN 443  
 DB 317 SPNFTINNETGDIITVAAGLDREKVQYTLIIQATDMEGNPTGLSNATATAVITVDVN 376  
 QY 444 DQIPF-EKSDYGNLTAEADTNIGSTILTIQATDADEFTGSSKILYHIKDGSEGRGV 502  
 DB 377 DNPPEFTAMTFYGEV---PENRVDIIIVANLTVDKQDQHTPAWNAVYRISGDEPTGFAI 433  
 QY 503 DTDPHNTGYVVIKKPLDDEFATAVSNIVFKAENPEPLVFGVKNASSFAKFTLIVTDVNE 562  
 DB 434 QTDPSNDGLTVVVKPIDFTNRMFVLTVAENQVPLAKIQHPPOSTATVSVTVIDVNE 493  
 QY 563 APQFSQHVFOAKVSEDAVIGTKVGNVAKDPEGL---DISYSLRGDTRGWLKIDHVTGEI 619

Db 494 NPYFAPNPKIIRQEBGLHAGTMLTFTTAQDDPRVMQONIRYTKLSDFANWLKIDPVNGQI 553  
Qy 620 FSVAPLDREA-----GSPYRVQVATEVGGSSLSVSEFHLILMDVNDNPPRLAKDYTGLE 675  
Db 554 TTIIVLDRESNVKNYINATFLASDNGGIPPMSCGTGLIYLLDINDNAPQVL----- 606  
Qy 676 FCHPLSA-----PGSLIFEATDDOHLFRGPH-FTFSLGSGSLQNDWEVSKINGTHAR 727  
Db 607 ---PQEAETCTPDPSNINITALDYIDPNAGPEAFDLPLSPVTKIKRNTTIRLNGDFAQ 663  
Qy 728 LSTRHTDEERAYVVLIRINDGRRPLGIVSLPVTFCSC-VEGSCFRPAGHGQIPIV- 785  
Db 664 LNLKIKFLEAGIYEYPIIITGSGPPKSNISILRVKVCOCDSNGDC-----TDVDRIV 716  
Qy 786 --GMAVGILLTLLVIGILAVFIRIKKDKGKNVESQAQSEVKP 829  
Db 717 GAGLGTGAIALLCIIILLVLMFVVMKRRDKERQAKQLLIDP 762

RESULT 10  
CAD2\_CHICK STANDARD; PRT; 912 AA.  
AC P10288; Q90630;  
AT 01-MAR-1989 (Rel. 10, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Neural-cadherin precursor (N-cadherin) (Cadherin-2).  
GN CDH2.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88153917; PubMed=2831236;  
RA Hatta K., Nose A., Nagafuchi A., Takeichi M.;  
RT "Cloning and expression of cDNA encoding a neural calcium-dependent  
cell adhesion molecule: its identity in the cadherin gene family.";  
J. Cell Biol. 106:873-881(1988).  
RN [2]  
RP SEQUENCE OF 1-25 FROM N.A.  
RC STRAIN=Cornish white rock Cockerel;  
RX MEDLINE=97354288; PubMed=9210582;  
RA Li B., Paradies N.E., Brackenbury R.W.;  
RT "Isolation and characterization of the promoter region of the chicken  
N-cadherin gene.";  
Gene 191:7-13(1997).  
CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.  
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC  
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE  
CC SORTING OF HETEROGENEOUS CELL TYPES. N-CADHERIN MAY BE INVOLVED IN  
CC NEURONAL RECOGNITION MECHANISM.  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X07277; CAA30258.1; -  
DR EMBL; U15563; AAB62980.1; -  
DR PIR; A29964; IJCHCN.  
DR HSP; P15116; INCUJ.  
DR InterPro; IPR002126; Cadherin.  
DR InterPro; IPR000233; Cadherin\_C\_term.  
DR Pfam; PF00028; cadherin; 5.  
DR Pfam; PF01049; cadherin\_C\_term; 1.  
DR PRINTS; PR00205; CADHERIN.

DR SMART; SM00112; CA; 5.  
DR PROSITE; PS00232; CADHERIN\_1; 3.  
DR PROSITE; PS00268; CADHERIN\_2; 5.  
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;  
Signal.  
FT SIGNAL 1 28 POTENTIAL.  
FT PROPEP 29 164 POTENTIAL.  
FT CHAIN 165 912 NEURAL-CADHERIN.  
FT DOMAIN 165 729 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 730 752 POTENTIAL.  
FT CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 753 912 CADHERIN 1.  
FT DOMAIN 165 272 CADHERIN 2.  
FT DOMAIN 273 387 CADHERIN 3.  
FT DOMAIN 388 502 CADHERIN 4.  
FT DOMAIN 503 609 CADHERIN 5.  
FT DOMAIN 610 720 SER-RICH.  
FT DOMAIN 869 884 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 278 278 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 407 407 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 628 628 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 657 657 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 21 21 A -> G (IN REF. 2).  
SQ SEQUENCE 912 AA; 100464 MW; 9BA5AC9DC1FFC489 CRC64;

Query Match 16.1%; Score 696; DB 1; Length 912;  
Best Local Similarity 26.9%; Pred. No. 6.7e-35;  
Matches 209; Conservative 117; Mismatches 328; Indels 124; Gaps 22;  
Qy 126 FLOSKYEGSVQRNSRPGKPFYLVNATDLDPATPNQQLYQIVQLPMINNMVYQIN-- 183  
Db 41 PPDVHSAVSRSHVGGQPLNVRFSQDSE-----NRKIYGSSEP 81  
Qy 184 -----NKTGALSITREGSQELNPAKNPSYLNVISVKDMGGOS-----ENFSDDT 228  
Db 82 EDFRVEDGVVAERSFQLSAEPT-----FVVSARDKETOEWMQVKLTPEPAFTGAS 136  
Qy 229 SVDIIVTENI-----WKAPKPVEMVENSTDPHPKIITQVRNDFGA 269  
Db 137 EKQKKKIEDIIFPMQOYKSSHLKQKRDWVIP-PINLPENSRGPPFQELVIR----- 189  
Qy 270 QYSLVDKELKPRFESI-----DQ-----EGDIYVTPQLDREKDAYVYAVAK 313  
Db 190 -----SDRKSLSLRYSVTGPGADPPTGFIINPISGQLSVTKPLDREQIASFHLRAHAV 245  
Qy 314 DEYGRPLSYPLEIHVKVDINDNPTCPSPVTVFEQENERLGNISIGTLTAHDRENTA 373  
Db 246 DVNGNQVENPIDIVINVIDMNDNRPEFLHQVWNGTVPEGSKPGTYVMVTALDADPNQA 305  
Qy 374 NSFLNYRIVEQTPKLPMDGLFLIQTAYAGMLQAKSLKKQDTPQYNLTIEVSDKD----- 428  
Db 306 NGMLRYILSQAPSSPSPNMFTINNETGDIITVAAGLOREKVQOYTLIIQATDMGNPTY 365  
Qy 429 -FKTLCFQVINVIDINDQIPIF-EKSDYGNLTLAEDTNIIGTILATQATDADPEPTGSSK 486  
Db 366 GLSNATATAVITVDNDNPPETAMTFYGEV---PENRVDVIVANLTVTDKQPHTPAWN 422  
Qy 487 ILYHIKGDSEGRGLVDTPHTNTGYVIKKPLDDETAAVSNIVKAEPELPGVKYN 546  
Db 423 ARYQMTGGDPTQGTILTDPSNNDGLVTWKPIDFETNRMFVLTVAAENQVPLAKGIQHP 482  
Qy 547 ASSFAKFLIVTDVNEAQFQSVQAKVSEDAVGTGKVGNTAKDP-----EGLDISYSL 602  
Db 483 POSTATVSTIVDVNESPYFVFNPKLVROEEGLLAGSMLTTFRTARDPRYMQOTSURYSK 542  
Qy 603 RGDTRGWLKIDHVTGEIFSVAPLDREA-----GSPYRVQVATEVGGSSLSVSEFHLILM 658  
Db 543 LSDPANWLKIDPVNGQIITTAVLDRSEIYVQNNMYNATFLASDNGIIPPMSCGTGLIYLL 602  
Qy 659 DVNDNPPRL-AKDYTGFLFCHPLSPAGSLIFEAETDDOHLFRGPHPTFSLGSG--SLQND 715  
Db 603 DINDNAPQVNPKEAT---TCETLQ-PNAINITAVDPDIDPNAGP-PAFELPDPSPPSIKRN 657

Qy 716 MEVSKINGTHARSTRHTRDFFERAYVVLIRINDGRPFLEGIVSLPVTFCS-VEGSCFR 774  
 Db 658 WTVIRISGHAQLSLRIFLEAGYDVPVITDSGNPHASSTSVLKVKVQCQDINGDC-- 715  
 Qy 775 PAGHOTGIPV---CMVAGILLTLLVIGILAVFIRIKKDKGKDNVESQAQSEVKP 829  
 Db 716 -----TDVIRVAGLGTGAIIALLCIIILILVLMFVVMKRRDKERQAKQLLIDP 768  
 RESULT 11  
 CAD2\_MOUSE  
 ID CAD2\_MOUSE STANDARD; PRT; 906 AA.  
 AC P15116; Q64260;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Neural-cadherin precursor (N-cadherin) (Cadherin-2).  
 GN CDH2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89346748; PubMed=2762814;  
 RA Miyatani S., Shimamura K., Hattori M., Nagafuchi A., Nose A.,  
 RA Matsunaga M., Hattori K., Takeichi M.;  
 RT "Neural cadherin: role in selective cell-cell adhesion";  
 RL Science 245:631-635(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Tamura K.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP PARTIAL SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6;  
 RX MEDLINE=92409532; PubMed=1528849;  
 RA Miyatani S., Copeland N.G., Gilbert D.J., Jenkins N.A., Takeichi M.;  
 RT "Genomic structure and chromosomal mapping of the mouse N-cadherin  
 gene";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:8443-8447(1992).  
 RN [4]  
 RP DEVELOPMENTAL STAGE.  
 RC STRAIN=C57BL/6; TISSUE=Testis;  
 RX MEDLINE=97033837; PubMed=8879495;  
 RA Munro S.B., Blaschuk O.W.;  
 RT "A comprehensive survey of the cadherins expressed in the testes of  
 fetal, immature, and adult mice utilizing the polymerase chain  
 reaction";  
 RL Biol. Reprod. 55:822-827(1996).  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 160-267.  
 RX MEDLINE=95191680; PubMed=7885471;  
 RA Shapiro L., Fannon A.M., Kwong P.D., Thompson A., Lehmann M.S.,  
 RA Grubel G., Legrand J.-F., Als-Nielsen J., Colman D.R.,  
 RA Hendrickson W.A.;  
 RT "Structural basis of cell-cell adhesion by cadherins";  
 RL Nature 374:327-337(1995).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (3.4 ANGSTROMS) OF 160-374.  
 RX MEDLINE=98318235; PubMed=9655503;  
 RA Tamura K., Shan W.S., Hendrickson W.A., Colman D.R., Shapiro L.;  
 RT "Structure-function analysis of cell adhesion by neural (N-)  
 cadherin";  
 RL Neuron 20:1153-1163(1998).  
 CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.  
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC  
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE  
 CC SORTING OF HETEROGENEOUS CELL TYPES. N-CADHERIN MAY BE INVOLVED IN  
 CC NEURONAL RECOGNITION MECHANISM.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED AT ALL STAGES OF TESTICULAR

CC DEVELOPMENT WITH HIGHEST LEVELS FOUND IN TESTES OF 21-DAY-OLD  
 CC MICE.  
 CC -!- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M31131; AAA37353.1; .  
 DR EMBL; AB008811; BAA23549.1; .  
 DR EMBL; S45011; AAB23356.1; .  
 DR PIR; A32759; IJMSCN.  
 DR PDB; INCG; 10-JUL-95.  
 DR PDB; INCH; 10-JUL-95.  
 DR PDB; INCI; 18-MAR-99.  
 DR MGD; MGI:88355; Cdh2.  
 DR InterPro; IPR002126; Cadherin.  
 DR InterPro; IPR000233; Cadherin\_C\_term.  
 DR Pfam; PF00028; cadherin; 5.  
 DR PRINTS; PR00205; CADHERIN.  
 DR SMART; SM00112; CA; 5.  
 DR PROSITE; PS00232; CADHERIN\_1; 3.  
 DR PROSITE; PS0268; CADHERIN\_2; 5.  
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;  
 KW Signal; 3D-structure.  
 FT SIGNAL 1 23  
 FT PROPEP 24 159  
 FT CHAIN 160 906  
 FT DOMAIN 160 724  
 FT TRANSMEM 725 746  
 FT DOMAIN 747 906  
 FT DOMAIN 160 267  
 FT DOMAIN 268 382  
 FT DOMAIN 383 497  
 FT DOMAIN 498 603  
 FT DOMAIN 604 717  
 FT DOMAIN 863 878  
 FT CARBOHYD 190 190  
 FT CARBOHYD 273 273  
 FT CARBOHYD 325 325  
 FT CARBOHYD 402 402  
 FT CARBOHYD 572 572  
 FT CARBOHYD 651 651  
 FT CARBOHYD 692 692  
 FT CONFLICT 7 9  
 FT CONFLICT 565 565  
 FT CONFLICT 567 567  
 FT CONFLICT 624 624  
 SQ SEQUENCE 906 AA; 99761 MW; 0206741B71707B7 CRC64;  
 Query Match 16.1%; Score 694.5; DB 1; Length 906;  
 Best Local Similarity 27.6%; Pred. No. 8.2e-35;  
 Matches 211; Conservative 122; Mismatches 332; Indels 99; Gaps 23;  
 Qy 117 KDINDNRPFTLQSKYGVSRQNSRFGKPLXVYNATDLPATPNQLYYQIVQLPMINN 176  
 Db 47 KDVHGQP-LLNVKFSNCNRK-----RKQVSESEPADKVDGDTVI--AVRSPFLTA 98  
 Qy 177 VMYFOI---NNKTG-----AISLRGQSEQLNPAKPN--SYNLVIVSKDM---GGQSENSF 224  
 Db 99 QAKFLIYAQDKETQKQVAVNLSREPTLTETPEKPEHEIEIVPRQLAKHSGALQRK 158  
 Qy 225 SDTTSVDIIVTENTKAPKPVENVSDPHPIKTVQRWN-----DPAQYS 272  
 Db 159 RD-----WVIP-PINLPENSGPPFQELVIRSDRKNLRLSVYVTGPGA--- 202  
 Qy 273 LVDRKELPRFPFSDIQ-EGDIYVTPQLDREKDAYVFYAVAKDEYKPLSVPLEIHKVK 331

```
Db 203 ----DQPTGFIINPISQSLSVTKPLDRELTARFHLRAHVAIDNGNQVENPIDIVINVI 258
Qy 332 DINDNPPPCSPVTVFVEQNERLGNSTGLTAHDDREENTANSLNRYIVVEOTPKLMD 391
Db 259 DMDNRPEFLHQVWNGSPVSGKPGTYVTVAIDADDPNALNGMLRYILSOAPSTSP 318
Qy 392 GLFLQTYAGMLQALQSLKKQDTPQYNLTIEVSKD-----FKTLCFVQINVIDNQ 445
Db 319 NMFTINNETGDIITVAAGLDREKVQYTLIIQATDMEGNPTYGLSNTATAVITVDVND 378
Qy 446 IPIF-EKSDYGNLTAEDTNGSTLTITQATDADPPFTGSSKILYHIKDGSEGLGVD 504
Db 379 PPEFTAMTFYGEV---PENRVDVIVANLTVTDKQDPHTPAWNAAYRISGGDPTGRFAIT 435
Qy 505 DPHNTGTVIILKPLDFFETAASNVIVFAENPEPLVFGVKYNASSFAKFTLLIVTDVNEAP 564
Db 436 DPNNDGLVTVVVKPIDFETNMFVLTVAENQVPLAKIQHPQSTAVSVTVIVDVENP 495
Qy 565 QFSQHVQAKVSEDVAICTKGVNVTAKDPEGL---DISYSLRGDTRGWLKIDHVTGEIFS 621
Db 496 VFAPNPKIIRQEBGLHAGTMTLTTLTAQDPDRYMQONIRYTKLSDPANWLKIDPVNGQITT 555
Qy 622 VAPLDREA---GSPYRQVQVATEVGGSSLSVSEFHLILMDVNDNPPRLAKDYTGLEFC 677
Db 556 IAVLDRESYPVQNNYNTATFLASDNGIPPMSTGTGLQIYLLDINDNAPQVL----- 606
Qy 678 HPLSA-----PGSLIFEATDDQHLFRGPH-FTFSLGSGSLONDWEVSKINGTHARLS 729
Db 607 -POEATCETPEPNSINTAALDYDIDPNAGPFADPLSPVYIKRWNINRLNGDFAQLN 665
Qy 730 TRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFSC-VEGSCFRPAGHQGTPTV--- 785
Db 666 LKIKLEAGIYEVPIIITDSGNPPKSNISILRVKVCQDSNGDC-----TDVDRIVGA 718
Qy 786 GNAVGLITLLVIGIILAVFIRIKKDKGKNVESAQAQSEVKP 829
Db 719 GLGTGAIITALLCIILLILVLMFVWMKRRDKERQAKQLLIDP 762

RESULT 12
CAD2_BOVIN
ID CAD2_BOVIN STANDARD; PRT; 877 AA.
AC P19534;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neural-cadherin precursor (N-cadherin) (Cadherin-2) (Fragment).
GN CDH2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90360979; PubMed=2390969;
RA Liaw C.W., Cannon C., Power M.D., Kibonaka P.K., Rubin L.L.;
RT "Identification and cloning of two species of cadherins in bovine
RT endothelial cells.";
RL EMBO J. 9:2701-2708(1990).
CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES. N-CADHERIN MAY BE INVOLVED IN
CC NEURONAL RECOGNITION MECHANISM.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X53615; CAA37677.1; -.
DR PIR: S11693; IJBOCN.
DR HSP: P15116; INCI.
DR InterPro: IPR002126; Cadherin.
DR InterPro: IPR000233; Cadherin_C_term.
DR Pfam: PF00028; cadherin; 5.
DR Pfam: PF01049; Cadherin_C_term; 1.
DR PRINTS: PR00205; CADHERIN.
DR SMART: SM00112; CA; 5.
DR PROSITE: PS00232; CADHERIN_1; 3.
DR PROSITE: PS0268; CADHERIN_2; 5.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
FT NON_TER 1
FT PROPEP <1 130
FT CHAIN 131 877
FT DOMAIN 131 695
FT TRANSEM 696 717
FT DOMAIN 718 877
FT DOMAIN 131 238
FT DOMAIN 239 363
FT DOMAIN 364 468
FT DOMAIN 469 574
FT DOMAIN 575 685
FT DOMAIN 834 849
FT CARBOHYD 161 161
FT CARBOHYD 244 244
FT CARBOHYD 296 296
FT CARBOHYD 373 373
FT CARBOHYD 543 543
FT CARBOHYD 622 622
FT CARBOHYD 663 663
SQ SEQUENCE 877 AA; 96845 MW; 441B829ED871A249 CRC64;
Query Match 15.9%; Score 688.5; DB 1; Length 877;
Best Local Similarity 26.7%; Pred. No. 1.8e-34;
Matches 203; Conservative 126; Mismatches 343; Indels 87; Gaps 19;
Qy 126 FLQSKYEGSVRONSRPGKPFVYNATDLDPATPGQYQIVQIQLP-----MINN 176
Db 7 FPDVYSAVLSDRDVLEGGQLLVNKFSC-----NGKRKVQYSESEPADEFKVDDEGMVYA 60
Qy 177 VNYFOINNKTAISLTREGSQELNPAKNPSYNLVIVSKDMGQSQENSEFSDTTSVDIIV-- 234
Db 61 VRSFPLSSSEHSLIYAQDK-----TOEKQWQVAVKLSLKPALPEDSVKRESREIEIVFP 115
Qy 235 -----TENIWKAPKPVEMVNSTDPHPKITQVRWN-----DPGAQ 270
Db 116 RQVTHNGYLQQRKRDWVIP-PINLPENSRGPPQELVIRSDRDNKNSLRSYVTPGPA- 173
Qy 271 YSLVDKEKLPRFPESIDQ-EGDIYVTOPLDREKDAYVYAVAKDEYKPLSVPLRIHWK 329
Db 174 -----DQPTGFIINPISQSLSVTKPLDRELTARFHLRAHVAIDNGNQVENPIDIVIN 227
Qy 330 VKDINDNPPTCSPVTVFVEQNERLGNSTGLTAHDDREENTANSLNRYIVVEOTPKLP 389
Db 228 VIDMNDNRPEFLHQVWNGTVPEGSKPGTYVTVAIDADDPNALNGMLRYILSOAPSTP 287
Qy 390 MDGLFIQTYAGMLQALQSLKKQDTPQYNLTIEVSKD-----FKTLCFVQINVIDIN 443
Db 288 SPNMTINNETGDIITVAAGLDREKVQYTLIIQATDMEGNPTYGLSNTATAVITVDV 347
Qy 444 DOIPIF-EKSDYGNLTAEDTNGSTLTITQATDADPPFTGSSKILYHIKDGSEGLGV 502
Db 348 DNPPEFTAMTFYGEV---PENRVDVIVANLTVTDKQDPHTPAWNAAYRISGGDPAQFAI 404
Qy 503 DTPDHTNTGYIILKPLDFFETAASNVIVFAENPEPLVFGVKYNASSFAKFTLLIVTDVNE 562
Db 405 QTDPNSNDGLVTVVVKPIDFETNMFVLTVAENQVPLAKIQHPQSTAVSVTVIVDNE 464
```

```
QY 563 APOFSOHVQAKVSEDAVIGTKVGNVYAKDPEGL---DISYSLRGDTRGWLKIDHVTGEI 619
Db 465 NPYFAENPKLIIROEGLHAGTLLTFTAQDPDRYMOQNIIRYTKLSDPANWLKIDSVNGQI 524
QY 620 FSVAPLDREA---GSPYRVQVATEVGGSSLSVSEFHLILMDVNDNPRLAKDYTGLF 675
Db 525 TTIADVLRSPNVKANIYNATFLASDNGIPPSGTTGLQYLDDINDNAPQVLPQBAEI- 583
QY 676 FCHPLSAPGSLIFEATDDQHILFRGPH-FTFSLGSGSLQNDWEVSKINGTHARLSRHTD 734
Db 584 -CE-TDPNSINITALDYIDPNAGPFAFDLPSPVTIKRNWTITRLNGDPAQLNLIKF 641
QY 735 FEERAYVILIRINDGRPPLEGIVLPTFCSC-VEGSCFRPAGHOTGPTV---GMAVG 790
Db 642 LEAGIYEVPIIIITDSGNPKSNISILIRVKVQCDSNGDC-----TDVDRIVGAGLGTG 694
QY 791 ILLTLLVIGIILAVVFIIRIKDKGKDNVESQAASEVKP 829
Db 695 AIIAILLCIILLILVMEVVMKRRDKERQAKQLLIDP 733

RESULT 13
CAD2_RAT
ID CAD2_RAT STANDARD; PRT; 906 AA.
AC Q921Y3; Q9R0T5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neural-cadherin precursor (N-cadherin) (Cadherin-2).
GN CDH2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC STRAIN-Sprague-Dawley; TISSUE=Testis;
RX MEDLINE=98187820; PubMed=9528971;
RA Chung S.S., Mo M.Y., Silvestrini B., Lee W.M., Cheng C.Y.;
RT "Rat testicular N-cadherin: its complementary deoxyribonucleic acid
RT cloning and regulation.";
RL Endocrinology 139:1853-1862(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-Sprague-Dawley; TISSUE=Brain;
RA Asai K., Tada T., Yamamoto M., Tada A., Mizuno M., Eimoto T., Kato T.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC !- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES. N-CADHERIN MAY BE INVOLVED IN
CC NEURONAL RECOGNITION MECHANISM.
CC !- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC !- TISSUE SPECIFICITY: IN TESTIS, EXPRESSED IN SERTOLI AND GERM
CC CELLS.
CC !- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF097593; AAC83818.1;
CC DB: EMBL; AB017695; BAA84919.1;
CC DR: HSSP; P15116; INCB.
CC
CC InterPro: IPR002126; Cadherin.
CC DR: Pfam; PF000233; Cadherin_C_term.
CC DR: Pfam; PF01049; cadherin; 5.
CC DR: PRINTS; PR00205; CADHERIN.
```

```
DR SMART; SM00112; CA; 5.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS00268; CADHERIN_2; 5.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW Signal.
FT SIGNAL 1 27 POTENTIAL.
FT PROPEP 28 159 POTENTIAL.
FT CHAIN 160 906 NEURAL-CADHERIN.
FT DOMAIN 160 724 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 725 745 POTENTIAL.
FT DOMAIN 746 906 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 160 267 CADHERIN 1.
FT DOMAIN 268 382 CADHERIN 2.
FT DOMAIN 383 497 CADHERIN 3.
FT DOMAIN 498 603 CADHERIN 4.
FT DOMAIN 604 717 CADHERIN 5.
FT DOMAIN 863 878 SER-RICH.
FT CARBOHYD 190 190 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 273 273 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 325 325 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 402 402 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 572 572 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 622 622 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 651 651 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 692 692 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 7 7 G -> A (IN REF. 2).
FT CONFLICT 48 48 T -> D (IN REF. 2).
FT CONFLICT 153 153 A -> R (IN REF. 2).
FT CONFLICT 646 646 A -> V (IN REF. 2).
FT CONFLICT 658 658 N -> K (IN REF. 2).
FT CONFLICT 724 724 T -> A (IN REF. 2).
SQ SEQUENCE 906 AA; 99685 MW; 97D9937FD8D1F8B5 CRC64;

Query Match 15.9%; Score 686.5; DB 1; Length 906;
Best Local Similarity 27.2%; Pred. No. 2.5e-34;
Matches 215; Conservative 126; Mismatches 344; Indels 105; Gaps 24;

QY 97 AALDANGIIV-----EGPVPTITIEVKDINDNRPITLQSKYEGSVQRNSRPGKPFLLYVNA 150
Db 21 ASLEASGELALCKTGFPEDVYSAVLPKTVHEGQP-LLNVKFSNCNRK----RKVQYES 74
QY 151 TDLDDPATNGQLYYQIVQLPMINVMYFQI---NNKYG-----AISLTREGSQELNPA 202
Db 75 EPADEKVEDGTVY--AVRSFPLSABQAKFLIYAQDKQEKQKQVAVNLSLEPSTEEM 132
QY 203 KNP-SYNLVISVKDM---GGQSENSFSDTSDIIVTENIWKAPKPVEMVENSSTDHPK 258
Db 133 KEPHEIEEIVFPQLAKHSGALQQRKD-----WVIP-PINLPENSRGFPFQE 179
QY 259 ITQVRWN-----DPGAQYSLVDKEKLPREPFSIDQ-EGDIYVYTPQLDRBEKDA 305
Db 180 LVRIISDRDKNLSRLSVYTGPGA-----DQPTGFIINPISGQLSVTKPLDRELIA 232
QY 306 YFYFAVAKDEYKPLSYPLEIHVKVKDINDNPTCSPVTVFEVQENRGLNSIGILTAAH 365
Db 233 FHLRAHAVDINGNQVENPDIVINVDMDNRPFEFLHQVWNGSVPEGSKPGTVYMTVTAI 292
QY 366 DRDEENTANSFNYRIVEGTPKLPMDGLFLIQYAGMLQALQSLKKQDPQYNLTIEVS 425
Db 293 DADDPNALNGMLRYRLSAPSPSPNMFTINNETGDIITVAAGLDREKVOQYTLIQAT 352
QY 426 DKD-----FKTLCFVQINVIDINDQIPF-EKSDYGNLTLABDTNIGSTILTQATDAD 478
Db 353 DMEGNPTYGLSNATATATVITVDVNDNPPETAMTFGEV---PENRVDVIVANLTVTDKD 409
QY 479 EPTGSSKLYLHIKDSGRGLVDTPDPTNTGTYVIKPKPLDPEETAASVNIKAEKNPEP 538
Db 410 QPHTPAWNAAYRISGGDPTGREFAILDTPNSNDGLVTVVWPKPDIFFTMRMFLVTAENQVP 469
QY 539 LVFGVKYNASSPAKFTLIYTDVNEAPQFQSOHVFAQVSESDVAITGKVGNNVAKDPGL-- 596
Db 470 LAKGIOHPQOSTATVSVTVIDVNNENPYFAPNPKIIRQEEGLHAGTLLTTLTAODPDYMQ 529
```

QY 597 -DISYSLRGDTRGWLKIDHTVGEIFSVAPLDREA---GSPYRVQVYVATEVGGSSLSVS 651  
 Db 530 QNIRITKLSDPANWLKIDPVNGQITITIAVLDRSPNKNYINATFLASDNGIPPMGSGTG 589  
 QY 652 EFHILMDVNDNPPRLAKDYTGFLFCFCHPLSA-----PGSLIFEATDDQHLFRGPH-F 703  
 Db 590 TLQIYLLDINDNAQVL-----PQEAETCETEPNSINITALDYDIDPNAGPFAF 639  
 QY 704 TFSLGSGLQNDWEVSKINGTHARLSTRHTDFFERAYVVLIRINDGRRPPLLEGIVSLPVT 763  
 Db 640 DLPLSPATIKRNTWITRLNGDFAQLNLKIKLEAGIYEVPIVITDSGNPPKSNISILRVK 699  
 QY 764 FCSC-VEGSCFRPAGHGTPTV---GMAYCILLTLLVIGIILAVFIRIKKDKGDNV 819  
 Db 700 VCQDSNGDC-----TDVRIVGAGLGTGTITAILLIIILILVLMFVVMKRRDKE 752  
 QY 820 ESAQASEVKP 829  
 Db 753 ROAKOLLIDP 762

RESULT 14  
 ID CAD4\_HUMAN STANDARD; PRT; 916 AA.  
 AC P5283;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Cadherin-4 precursor (Retinal-cadherin) (R-cadherin) (R-CAD).  
 GN CDH4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=95073006; PubMed=7982033;  
 RA Tanihara H., Sano K., Helmark R.L., St John T., Suzuki S.;  
 RT "Cloning of five human cadherins clarifies characteristic features of  
 RT cadherin extracellular domain and provides further evidence for two  
 RT structurally different types of cadherin.";  
 RL Cell Adhes. Commun. 2:15-26(1994).  
 RN [2]  
 RP SEQUENCE OF 393-916 FROM N.A.  
 RC TISSUE=Fetal brain;  
 RX MEDLINE=91283540; PubMed=2059658;  
 RA Suzuki S., Sano K., Tanihara H.;  
 RT "Diversity of the cadherin family: evidence for eight new cadherins  
 RT in nervous tissue.";  
 RL Cell Regul. 2:261-270(1991).  
 CC -|- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.  
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC  
 CC MANNER IN CONNECTING CELLS. CADHERINS MAY THUS CONTRIBUTE TO THE  
 CC SORTING OF HETEROGENEOUS CELL TYPES. MAY PLAY AN IMPORTANT ROLE  
 CC IN RETINAL DEVELOPMENT.  
 CC -|- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -|- TISSUE SPECIFICITY: EXPRESSED MAINLY IN BRAIN BUT ALSO FOUND IN  
 CC OTHER TISSUES.  
 CC -|- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: L34059; AAA35627.1; -  
 CC DR HSSP: P15116; INCI.  
 CC DR Genew: HGNC:1763; CDH4.  
 CC DR MIM: 603006; -

DR InterPro; IPR002126; Cadherin.  
 DR InterPro; IPR000233; Cadherin\_C\_term.  
 DR Pfam; PF00028; cadherin; 5.  
 DR Pfam; PF01049; Cadherin\_C\_term; 1.  
 DR PRINTS; PR00205; CADHERIN.  
 DR SMART; SM00112; CA; 5.  
 DR PROSITE; PS00232; CADHERIN\_1; 2.  
 DR PROSITE; PS0268; CADHERIN\_2; 5.  
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;  
 KW Signal.  
 FT SIGNAL 1 20 POTENTIAL.  
 FT PROPEP 21 169 POTENTIAL.  
 FT CHAIN 170 916 CADHERIN-4.  
 FT DOMAIN 170 734 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 735 756 POTENTIAL.  
 FT DOMAIN 757 916 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 777 916 CADHERIN 1.  
 FT DOMAIN 278 392 CADHERIN 2.  
 FT DOMAIN 393 507 CADHERIN 3.  
 FT DOMAIN 508 613 CADHERIN 4.  
 FT DOMAIN 614 724 CADHERIN 5.  
 FT DOMAIN 725 888 SER-RICH.  
 FT CARBOHYD 283 283 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 412 412 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 557 557 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 632 632 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 661 661 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 702 702 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 916 AA; 100446 MW; E717C54A19E0C52A CRC64;  
 Query Match 15.8%; Score 684.5; DB 1; Length 916;  
 Best Local Similarity 29.3%; Pred. No. 3.4e-34;  
 Matches 176; Conservative 106; Mismatches 288; Indels 31; Gaps 12;  
 QY 239 WKAPKPVEMVENSDDPHIKITQVRW---NDPGAQYSL---VDREKLPFRFESID-QSGDI 292  
 Db 171 WVIP-PINVPENSRGPPQQLVRIIRSKDNDIPRYSITGVGADQPPMEVFINSMSGRM 229  
 QY 293 YVTQPLDREEDKDAYFYAVAKDEYVGKPLSYPLETHVKKVINDNDNPCTPSPVTVFVQEN 352  
 Db 230 YVTRPMDREEHASHYHLRAHADVNGKNVNPIDLYIYVIDMNDNHPEFIQVYNCVDEG 289  
 QY 353 ERLNGSIGTUTAHDRDEENTANSFLNRYVEQTPKLPMDGLFLIQTIVAGMLQIAQSLKK 412  
 Db 290 SKPGTYVMTITANDADDDSTTANGMVRYRIVTQTPQSPSQNMFTINSETGDIVTVAAGWDR 349  
 QY 413 QDPPOYNLTIEVSKD-----FKTLCFQVINVIDNQIPIEKSDYGNLTAEADNIG 466  
 Db 350 EKVOQTTVIVQATDMEGNLNYSNTATATITVTDVNDNPSEFTASTFAG--EVPENSE 407  
 QY 467 STILTIQATDADPEFTGSSKILYHIKGDSEGRGLVDTPDHTNTGYVVIKKPLDFETA 526  
 Db 408 TVVANLTVMRDQPHSPNNNAVRIISGDSGHFSVRTDPTVNEGVMVTVKADVDYELNRA 467  
 QY 527 SNIVFAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQESQHVQAKVSEDAVIGTKVG 586  
 Db 468 FMLTVMVSNQAPLASGIONSFQSTAGVTISIMDINEAPYPPSNHKLIRLEGVPPGTULT 527  
 QY 587 NYTAKDEGL---DISYSLRGDTRGWLKIDHTVGEIFSVAPLDREA---GSPYRVQVVA 639  
 Db 528 TFSVADDFRPMQOAVRYSKLSDPASWLNATNGQITTVAVLDRESLYTKNNYVETFLA 587  
 QY 640 TEVGGSSLSVSEFHLITLMDVNDNPPRLAKDYTGFLFCFCHPLSAPGSLIFEATDDQHLFR 699  
 Db 588 ADNGIPASGTGTLQIYLLDINDNAPPELLKPEAQICERPNLNA---INITAADADHPNI 644  
 QY 700 GPH-FTFSLGSGSLQNDWEVSKINGTHARLSTRHTDFFERAYVVLIRINDGRRPPLLEGIV 758  
 Db 645 GPVVFELPEVPAARVKNWITRLNGDYAQLSLRILYLEAGMYDVPVITVDSGNPPLSNTS 704  
 QY 759 SLPTVTFSCVVEGSCFRPAGHGTPTVGVNAVGTLLTLLVIGIILAVV--FIRIKDKGK 816  
 Db 705 IIKVKVCPDDNGDCTTIG---AQAAGLGTGAIVAILICILILLTLLVLLFVMMKRRREK 761

QY 817 D 817  
Db 762 E 762

RESULT 15  
DSC2\_MOUSE  
ID DSC2\_MOUSE STANDARD; PRT; 902 AA.  
AC P55292; Q64734;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Desmocollin 2A/2B precursor (Epithelial type 2 desmocollin).  
GN DSC2 OR DSC3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Embryo;  
RX MEDLINE=95227276; PubMed=7711832;  
RA Lorimer J.E., Hall L.S., Clarke J.P., Collins J.E., Fleming T.P.,  
RA Garrod D.R.;  
RT "Cloning, sequence analysis and expression pattern of mouse  
RT desmocollin 2 (DSC2), a cadherin-like adhesion molecule.";  
RL Mol. Membr. Biol. 11:229-236(1994).  
RN [2]  
RP SEQUENCE OF 344-637 FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Embryo;  
RX MEDLINE=95048328; PubMed=7959727;  
RA Buxton R.S., Wheeler G.N., Pidsley S.C., Marsden M.D., Adams M.J.,  
RA Jenkins N.A., Gilbert D.J., Copeland N.G.;  
RT "Mouse desmocollin (Dsc3) and desmoglein (Dsgl) genes are closely  
RT linked in the proximal region of chromosome 18.";  
RL Genomics 21:510-516(1994).  
CC -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS. INVOLVED  
CC IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS  
CC MEDIATING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL CELL  
CC POSITIONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL  
CC ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOFORMS.  
CC -!- SUBCELLULAR LOCATION: Type 1 membrane protein.  
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 2A (SHOWN HERE) AND 2B; ARE  
CC PRODUCED BY ALTERNATIVE SPLICING.  
CC -!- TISSUE SPECIFICITY: IN ALL EPITHELIA TESTED AND HEART.  
CC -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS  
CC (POTENTIAL).  
CC -!- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: L33779; AAA79177.1; -  
CC EMBL: L33779; AAA79176.1; -  
CC EMBL: X73885; CAA52089.1; -  
CC HSP: P15116; INCJ.  
CC MGD: MGI:103221; Dsc2.  
CC InterPro: IPR002126; Cadherin.  
CC InterPro: IPR002123; Cadherin\_C\_term.  
CC Pfam: PF00028; cadherin; 5.  
CC Pfam: PF01049; Cadherin\_C\_term; 1.  
CC PRINTS: PR00205; CADHERIN.  
CC SMART: SM00112; CA; 5.  
CC PROSITE: PS00232; CADHERIN.1; 3.  
CC PROSITE: PS00268; CADHERIN.2; 5.  
CC Cell adhesion; Glycoprotein; Transmembrane; Repeat; Signal;  
KW Alternative splicing; Cytoskeleton; Calcium-binding.

FT	SIGNAL	1	27	POTENTIAL.
FT	PROPEP	28	135	POTENTIAL.
FT	CHAIN	136	902	DESMOCOLLIN 2A/2B.
FT	DOMAIN	136	694	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	695	715	POTENTIAL.
FT	DOMAIN	716	902	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	136	243	CADHERIN 1.
FT	DOMAIN	244	355	CADHERIN 2.
FT	DOMAIN	356	471	CADHERIN 3.
FT	DOMAIN	472	579	CADHERIN 4.
FT	DOMAIN	580	694	CADHERIN 5.
FT	CARBOHYD	166	166	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	392	392	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	546	546	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	629	629	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	838	848	KVQFCHTDNQ -> ETIRGHTLIKN (IN ISOFORM 2B).
FT	VARSPLIC	849	902	MISSING (IN ISOFORM 2B).
FT	SEQUENCE	902 AA;	99961 MW;	E233F8CB20ACCAB0 CRC64;

Query Match 15.6%; Score 674.5; DB 1; Length 902;  
Best Local Similarity 31.2%; Pred. No. 1.4e-33;  
Matches 185; Conservative 99; Mismatches 268; Indels 41; Gaps 17;

QY	239	WRAPKPVEMVNSTDPHPKIKITQVRWNDPGAQYSL-----VDKEKLPRFFSIDQEG	290
Db	137	W-APIPCSMLENSLGPFLFLOIQ-SDTAQNVTIYYSIRGPGVDKPLNLFYVERD-TG	193
QY	291	DIYVTPDLREKDAYVAVAKDEYKPLSYPLHIVKVKDINDNPPTCPSPVTFEVQ	350
Db	194	NLYCTGRVDREQYESFELTAFATTPDGYTEYPLPLIKIEDENDNYPITQKLYSTVQ	253
QY	351	ENERLGNSTGLTAHDRDEENTANFLNRYIVEQTPLKPMDFLFIQTYAGMLQAKSL	410
Db	254	ENSRIGSIVGEVCAITDLDPDTHMTRLYSILEQSPPM--LFTMHPSTGVITTTSAQL	311
QY	411	KKODTPOYNLTIEVSKD-----FKTLCFVOINVIDNDQIPEKSDYGNLTLAETNI	465
Db	312	DRELIDKQLLIRKQVMDGQYFGLHTTAKCIITIEDVNDNLPTFTRTTY--VTSVEENTV	369
QY	466	GSTILTQATDADEPTGSSKILYHIKDGSEGLRGVDTDPHTNTGVYIILKPLDFTAA	525
Db	370	NVEILRLTVQDKDLVNSPNRANRYITLKGNEGNFKIVTDPKTNNEGILCVIKPLDYERQ	429
QY	526	VSNIVEKAENPEPLVFCVKYNA--SSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTK	584
Db	430	QVTLQIGVVNEAPYTREASSKSPMSTATVTVTVTNQDEGPECIPPMQTVRIQENVPVGR	489
QY	585	VGNVTKADPE---GLDISYSLRGDTRGWLKIDHVTGTFISVAPLDREA----GSPYRVQV	637
Db	490	NDGYKAYDPETRSSSGIRYKLSDPGRGVTVNEDSGSITIFRALDREAETVRNGIYNTV	549
QY	638	VATEVGSSLSSEVSEHLLIMDVNDNPPRLAKDYTGFLFCHPLSPAGSLIFEATDDOHL	697
Db	550	LALDADGRSGTGT--LGIILEDVNDNGPFIPKO--TVVICKATMSSAEIVAVLDDP---	602
QY	698	FRGPHPTFSLGS--GSLQNDWEYSKINGTHARLSTRHTDPEERAYVYVLIIRINDGGRPLE	755
Db	603	VNGPPDFSLSSDSEVQRMRLTRINDTAARLSYQN-DPSFGSYAVPIRVTD--RLGLS	659
QY	756	GIVSLPVTFCSCV-EGSCFRPAGHQGTPIVGMVAGILLTTLVIGILLAVVF	807
Db	660	SVTTLNVLVCDCTITESDCTLRSGERTGYADVRLGPMAILAILGIALLCILF	712

Search completed: April 2, 2003, 15:52:48  
Job time : 22.0959 secs

**THIS PAGE BLANK (USPTO)**



Result No.	Query No.	Score	Query %		Length	DB	ID	Description
			Match					
1	1	2332	54.0	578	11	Q63423	Q63423 rattus norv	
2	1	1427.5	33.0	868	13	Q90X63	Q90X63 brachydanio	
3	3	709.5	16.4	893	13	Q8UVQ7	Q8UVQ7 brachydanio	
4	4	699	15.2	922	13	P79883	P79883 xenopus. ma	
5	6	677	15.7	783	13	Q90275	Q90275 brachydanio	
6	5	671.5	15.5	839	4	Q9HAZ9	Q9haz9 homo sapien	
7	7	652	15.1	713	4	Q8TBX3	Q8tbx3 homo sapien	
8	6	652	15.1	714	1	Q8VDPK4	Q8vdk4 mus musculus	
9	9	651.5	15.1	814	6	Q77704	Q77704 canis famil	
10	6	650	15.0	714	11	Q8R490	Q8r490 rattus norv	
11	11	628	14.5	4589	11	Q9WU10	Q9wu10 rattus norv	
12	6	624	14.4	4587	11	Q9QXA3	Q9qxa3 mus musculus	
13	13	611	14.1	4555	11	Q8R508	Q8r508 rattus norv	
14	4	601.5	13.9	3579	5	Q9V5N8	Q9v5n8 drosophila	
15	15	591.5	13.7	4643	5	Q9VW71	Q9vw71 drosophila	
16	16	576	13.3	3014	4	Q9NVQ6	Q9nvq6 homo sapien	

```
Db 1 LTGETGILKIEKDLGLHARVLDRETRAVHHLRLAALDQSGAIVDGPVPIIIIEVKDIND 60
QY 122 NRPTFQSKYEGSVQRNRPCKPFLYVYNATDLDPATPNQGLYYQYIQLPMNNVMYFQ 181
Db 61 NRPTFLQTKYEGSVQRNRPCKPMYVYNATDLDPATPNQGLFYQYIQLPKNNVMYFQ 120
QY 182 INNKTAISLTRGSOQLNPAKPNYSNLVTSVKDMGQSGENSESDTTSVDIIVTENTWKA 241
Db 121 IDNKTAISLTRGSOQLDPIKPNYSNLVTSVKDMGQSGENSESDTTSVDITVRENTWKA 180
QY 242 PKPVMYENSTDPHPKITQVRWMDCAQYSLVDKELPRFPFSIDQEGDIYVTOPLDRE 301
Db 181 PEPVEIRENTDPHPKITQVQNEPVAHSLINKELPQFPFSIDQEGNIYVTOPLDRE 240
QY 302 EKDAYEYAVAKDEYGRPLSYPLEIHVKVDINDNPPTCPSPVTVFEVQENERLGNISGT 361
Db 241 EKDSHVEFATAKDENGKPLAYLEIRVKVIDINDNPPTCLSQVTVFEVQENEVLGSSIGI 300
QY 362 LTAHDRDEENTANSFLNRYIVEQTPKPLMDGLFLIQTYYAGMLQAKSLKKQDTPQYNLT 421
Db 301 FAAHMDDEANNISILMYRLVDQTPKPSDEFLIDEYGGKVLGKRSLSKKQDSPQYNLT 360
QY 422 IEVSDKDFKTLCFQVQINVIDIQPIFEKSDYCNLTAEQTNIGSTILTQATDADEPF 481
Db 361 VEVSIDDFKTLCSLQVNVINDIQPIFEKSDYGSKLSLSEDTALGSTILTIQATDDDEPF 420
QY 482 TGSKILYHIKDGSEGLRGVDPHTNTGYVIKKPLDPETAAVSNIVFKAENPEPLVF 541
Db 421 TGSKILYKIVQDTEGRLEVDPMTNTGYVKIRKPLDTEPETEPTSIVFKAENPEPLVN 480
QY 542 GVKYNASSEFAKFTLVTDVNEAFOFSOHVFOAKVSEDAVGTGVNVTAKDPEGLDISYS 601
Db 481 GIEFNASSFASFELTVADNVEVPFQOIFQANVSEDTAIGTKVGTVTAKDPEGLTVSYS 540
QY 602 LRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVA 639
Db 541 LKDKRGWLKIDSVTGDIESTAPLDRETSYRVQVVA 578

RESULT 2
Q90X63
ID Q90X63 PRELIMINARY; PRT; 868 AA.
AC Q90X63;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Cadherin-17.
GN CDH17.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Horsfield J., Ramachandran A., Reuter K., Lavallie E.,
RA Collins-Racie L., Crosier K., Crosier P. Sr.;
RT "zebrafish cadherin-17 is required for pronephric duct development.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
RC -I- SIMILARITY: CONTAINS 6 CADHERIN DOMAINS.
DR EMBL; AF428098; AAL29444.1; -.
DR InterPro; IPR002126; Cadherin.
DR Pfam; PF00028; cadherin; 6.
DR PROSITE; PS00232; CADHERIN_1; UNKNOWN_2.
DR PROSITE; PS02688; CADHERIN_2; 6.
KW Calcium-binding; Cell adhesion; Glycoprotein.
SQ SEQUENCE 868 AA; 96677 MW; 40C30844C836E990 CRC64;

Query Match 33.0%; Score 1427.5; DB 13; Length 868;
Best Local Similarity 36.5%; Pred. No. 2.5e-85;
Matches 320; Conservative 150; Mismatches 330; Indels 77; Gaps 16;

QY 13 LMLYLATGYG--QEGKFSGLPKWPTSIVBQEPSQIIQF-KANPPAVTFELTGETDNI 69
```

```
Db 9 LTLVSGHGDLEDK-KGPLIDTVLDPPEATPVYAFFKFTSAVEDVSSYRSGETEDK 67
QY 70 FVTEREGLLYNRALDRETRSTHNLQVAALDANGIIIEGVPPIETVKDINDNRPTFLQS 129
Db 68 IRTSSDGLWLEQPLEWSPEKKHLLIEALSEDGKTLDGPAQVVLQVIDVNNNPVPS 127
QY 130 KYEGSVQRNRPCKPFLYVYNATDLDPATPNQGLYYQYIQLPMNNVMYFQINNKTAI 189
Db 128 QYSGSIREHSPAGVPFVQVFASDADDPTENTQLRFSIVNQIPVIGQTFEFGINPNQI 187
QY 190 SLTREGSOEL----- 199
Db 188 FTTEEGAEFLKARPSVTSYRGEVGRSPDLVKKPFEDYCIKNNIALENNPFYKVERAER 247
QY 200 ---NPAKNYSNLVTSVKDMGQSGENSESDTTSVDIIVTENIKAPKPMVEMNSTDPHP 256
Db 248 RTVNVLDQPDYALIVRAEDLGGNAVNSLSTTRVNIAILQNLWSPGPITIRENTDEYP 307
QY 257 IKITQVRWMDGQAYSLVDKELPRFPFSIDQEGDIYVTOPLDREKDAYEYVYAVAKDEY 316
Db 308 MYLATMRANNPTALYRLEQKEIL-SFPFTINQDGIYVTGTPLDREKEMYTLVVIAEDQ 366
QY 317 GKPLSYPLEITHVKVKDINDNPPTCPSPVTVFEVQENERLGNISGTLTAAHRODEENTANSF 376
Db 367 GVELEKPEWELPVLDQENDNPPMCDE--ALFEVQEKPEVPGNSIGHLPAHONDKEGTUSSA 424
QY 377 LNTYRIVQOTPKPMQDGLFLIQTYYAGMLQAKSLKKQDTPQYNLTIEVSKDFTKLFVQ 436
Db 425 LTYTLRSQTPTKPSDKMFSIDPNTGELKVANQFQKQVQYELTFEVDVFTKCKAI 484
QY 437 INVIDINQIPIFEKSDYCNLTAEQTNIGSTILTQATDADEPFGSSKILYHIKDGSD 496
Db 485 IKVIDINDEIPIFEKNDYGYTSVPPELAEVGTTLNIKATDADDTGTGSSRVYEHITAGDP 544
QY 497 EGRGLVDTPHTNTGYVIKKPLDPEFAAVSNIVFKAENPEPLVFYGVKNASSFAKFTLI 556
Db 545 QNLFAIEVDEETGEGRVYIAQPLDYELQSVNKLKIDARNPEPLIAGVEYNDSSITSVIE 604
QY 557 VTDVNEAQPFSQHVFOAKVSEDAVGTGVNVTAKDPEGLDISYSLRGDTRGWLKIDHVT 616
Db 605 LVDVDEPPKFEVEGLNVNVPENITVGTLLMKAEEKDPGKTIKPKMEGDEHKWLELVDT 664
QY 617 GEIFSVAPLDREAGSPYRVQVVADEVGSSLSVSEFHLIMDVNDNPPRLAKDYTLGLFF 676
Db 665 GELTKAALDRETVDFHTLTITAYETEGSKMEAKMVDIHLQDVNDNYPKLQK--TQGF 722
QY 677 CHPLSAPGSLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEVSKINGKINGHARLSTRHTD 736
Db 723 CLQDMTP--LTLTAMDKADPY-GEPTFAISRKS--QNEIFKPVDTGSAKLILKKPSS 777
QY 737 ERAYVVLIRINDGGRPPLEGIV-SLPVTFCSVE-GSCF-RPAGHOTGIPTVGMAVGILL 793
Db 778 EQNVTPPINVLNAG---LGIQKFDVRCNCKLGYCYIEPASHSNKL-SMGSTIGILA 833
QY 794 TTLVIGIILAVFIRIKKDKGKNVESASQAEVKPL 830
Db 834 GVFGVIGLFLGICLYQIKK---KDKQRTAETGKTAM 867

RESULT 3
Q8UVQ7
ID Q8UVQ7 PRELIMINARY; PRT; 893 AA.
AC Q8UVQ7;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE N-cadherin.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
```



Db 231 IFGRNVTRPLDRERSSYHLRAHAVDINGNKVENPIDLSIYVIDMNDNRPEFSSPIENG 290  
 Qy 348 EVOENRNLGNSIGTLTAHNDREENTANSEFLNRYVEQTPKLPMDGLFLIQTIVAGMLQAK 407  
 Db 291 SVDEASKGTVVMTADHADDINTSGIVMRYMDQSPQSPSHDMFVIHSGKTGVINTVA 350  
 Qy 408 QSLKKQDTPQYNLTIEVSKD-----FKTLCFVQINVIDINDQIPIFEKSDYGNLTALAE 461  
 Db 351 AGLDREKVVQYTVVLTQATDMEGNLNHLGNSNTATITVADVNDNPPETTRKMF--IGEPV 408  
 Qy 462 DTNIGSTILTQATDADPEFTGSSKLYHIKGDSEGRGLGVDTPDHTNTGYVIAKKPLDF 521  
 Db 409 ENHVDVVANLTVVDRDQPYTSNNNAVEKIIISGDPDGHGTIKTDVTVNEGIVTVSKPDY 468  
 Qy 522 ETAAVSNIVFAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFAKVEDVAI 581  
 Db 469 EMSKVFLLIVMTVNOAPLASQISLQSTAAVTVSVDVNEAPYPPNRPNEPIRLEGESA 528  
 Qy 582 GTKVGNVTAKDPE---GLDISYSLRGDTRGWLKIDHTVGEIFSVAPLDREAGSP----- 632  
 Db 529 GRLLITFSVADPHDSMQVQLRYSKISDPANLWNTTNGQVSTTAVLDRE--SPFVKDDL 586  
 Qy 633 YRVQVATEVCGSSLVSEFHLIMDVNDNPPR-LAKDYGTLFFCHPLSAPGSLIFEAT 691  
 Db 587 YQAKFLATDNGNPPASGTGTLIIQLIDINDNAPELLPKDAQ---ICERPNGNG-INITAI 642  
 Qy 692 DDDQHLFRGPHFTFSLGS--GSLQNDVEVSKINGTHARLSTRHTDFFERAVVVLIRINDG 749  
 Db 643 DVDRKPSADP-FVELSPVPTIRNTWTHIRNSIYARLSQIGYLSGMDVDPVIVTDS 701  
 Qy 750 GRPPLIGIVSLVPTFCSCVEGSCFRPAGHQGTIPVGMAGVILTLTLLVIGIILAVV--F 807  
 Db 702 GNPLYNTSIKVKVCPDNGNDCTTIG---AVAAAGLGTGALISILICIILLISWVLLF 758  
 Qy 808 IRIKDKGKD 817  
 Db 759 VMMKREKE 768

RESULT 5  
 Q90275 PRELIMINARY; PRT: 783 AA.  
 AC Q90275;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Neural-cadherin precursor (N-cadherin).  
 GN CDH2 OR CDH2 OR ZCAD.  
 OS Brachydanio rerio (zebrafish) (Zebra danio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 OX NCBI\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.  
 RC TISSUE=EMBRYO;  
 RX MEDLINE=95178741; PubMed=7873785;  
 RA Bitzur S., Kam Z., Geiger B.;  
 RT "Structure and distribution of N-cadherin in developing zebrafish embryos: morphogenetic effects of ectopic over-expression.";  
 RL Dev. Dyn. 201:121-136(1994).  
 CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS. THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC MANNER IN CONNECTING CELLS. CADHERINS MAY THUS CONTRIBUTE TO THE SORTING OF HETEROGENEOUS CELL TYPES. N-CADHERIN MAY BE INVOLVED IN NEURONAL RECOGNITION MECHANISM.  
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -!- TISSUE SPECIFICITY: EXPRESSION IN THE EMBRYO IS INITIALLY FOUND IN ALL DEEP CELLS BUT LATER BECOMES RESTRICTED TO VARIOUS EPITHELIAL AND NEURONAL TISSUES. FOUND IN DISCRETE AREAS OF CELL-CELL ADHESION WHEN EPIBOLY IS NEARLY COMPLETED. EXPRESSED IN ADULT BRAIN, EYES AND TRUNK.  
 CC -!- DEVELOPMENTAL STAGE: EXPRESSION IS FIRST DETECTED IN THE EMBRYO

CC DURING GASTRULATION.  
 CC -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.  
 DR EMBL; X67648; CAA47890.1; -.  
 DR HSSP; P15116; INCI.  
 DR ZFN; ZDB-GENE-990415-171; cdh2.  
 DR InterPro; IPR002126; Cadherin.  
 DR InterPro; IPR000233; Cadherin\_C\_term.  
 DR Pfam; PF00028; cadherin; 5.  
 DR Pfam; PF01049; Cadherin\_C\_term; 1.  
 DR PRINTS; PR00205; CADHERIN.  
 DR SMART; SM00112; CA; 5.  
 DR PROSITE; PS00232; CADHERIN\_1; 2.  
 DR PROSITE; PS0268; CADHERIN\_2; 5.  
 DR PROSITE; PS0268; CADHERIN\_3; 5.  
 KW Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;  
 KW Calcium-binding; Repeat; Signal.  
 FT SIGNAL 1 ? POTENTIAL.  
 FT PROPEP ? 35 POTENTIAL.  
 FT CHAIN 36 783 NEURAL-CADHERIN.  
 FT DOMAIN 36 601 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 602 622 POTENTIAL.  
 FT DOMAIN 623 783 CYTOPLASMIC (POTENTIAL).  
 FT REPEAT 36 143 CADHERIN 1.  
 FT REPEAT 144 258 CADHERIN 2.  
 FT REPEAT 259 373 CADHERIN 3.  
 FT REPEAT 374 480 CADHERIN 4.  
 FT REPEAT 481 591 CADHERIN 5.  
 FT DOMAIN 740 755 SER-RICH.  
 FT CARBOHYD 149 149 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 201 201 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 278 278 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 331 331 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 449 449 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 499 499 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 528 528 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 569 569 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 783 AA; 87120 MW; B57ECEDF41CC1C5 CRC64;  
 Query Match 15.7%; Score 677; DB 13; Length 783;  
 Best Local Similarity 30.1%; Pred. No. 5e-36;  
 Matches 187; Conservative 98; Mismatches 288; Indels 48; Gaps 15;  
 QY 239 WKAPKPVEMVENSVDHPH---IKITQVRWNDPGAYSLV--DKEKLPREFSIDQ-BGDI 292  
 Db 37 WVLP-PVNVLENSKQPEELVKIQSDKSNLRYSVTGPQADQNTGLFIIDISGLL 95  
 QY 293 YVTOPLDREKDAYVYFAVAKDEYKPLSYPLEIHVVKDINDNPPTCPSPVTTFVQEEN 352  
 Db 96 SVTKPLDRHPIPNFLRAHAVDINGNQMNENPIDIINVIDMNDNRPEFTHQINGTVDG 155  
 QY 353 ERLGNSIGTLTAHNDREENTANSEFLNRYVEQTPKLPMDGLFLIQTIVAGMLQAKSLK 412  
 Db 156 AKPGTFVMTVTSQDKDDPNTANGMLRYKILSQTPESPSSNNFTINNKTKIITVAAGLDR 215  
 QY 413 QDTPQYNLTIEVSKD-----FKTLCFVQINVIDINDQIPIFEKSDYGNLTALAE 466  
 Db 216 EKVPQTLIIQATDMEGNPTYGLSNTATATVIRLVNDVNDNAPEFTRETTHGHEPENRVNVI 275  
 QY 467 STILTIQATDADPEFTGSSKLYHIKGDSEGRGLGVDTPDHTNTGYVIAKKPLDFETA 526  
 Db 276 VTNLTV--TDKDEPGTPANNNAVRIISGDPTRGFSIPTDPTVNEGLTVVVKPVDFFEM 333  
 QY 527 SNIVFAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFAKVEDVAITKVG 586  
 Db 334 FMLTVVADVADNEVPLASGTHIRTRQSTATVSRVIDNESPNFDPNPKQIKLEGLPQWSMLT 393  
 QY 587 NYTA-KDPEGL---DLSYSLRGDTRGWLKIDHTVGEIFSVAPLDREA---GSPYRVQV 638  
 Db 394 TTAHDPPRYMOOTIYSKLYDPANLLEIDPNNGRISTIAVLDRSPYVKNLYNATFM 453  
 QY 639 ATEVGGSSLSVSEFHLIMDVNDNPPRLAKDYTGTLFFCHPLSA-----PGSLIFEAT 691  
 Db 454 ASDNGVPRASGTGTLIIQLIDINDNAPRV-----FPQEAECVCEPPEPNAINITAV 503

```

QY 692 DDQHFLRGPHTFTSLGS--GSLQNDWEVSKINGTHARLSRTHRTDPERAYVVLIRINDG 749
Db 504 DGLNPNAGP-YAFELPNRPSDIRRNTWLTTRISGDHAQLSKTISYSGIYELPISITDS 562
QY 750 GRPLBGIIVSLPTFCSC-VEGSCFRPAGHOTGIPVGNAGVGLLTLLVIGIILAVVFI 808
Db 563 GNLPMNSNTYLRKIVQCDDHGDV----DMERIMAAGLTGATIAILICIILLVLVLM 618
QY 809 RIKKDKGDNVESQAQSEVKP 829
Db 619 FYWMKRRDKERQAKQLLIDP 639

RESULT 6
Q9HAZ9
ID Q9HAZ9 PRELIMINARY; PRT; 839 AA.
AC Q9HAZ9
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Desmocollin 3b.
GN DSC3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20484176; PubMed=11027496;
RA Whitlock N.V., Hunt D.M., Rickman L., Malhi S., Vogazianou A.P.,
RA Dawson L.F., Eady R.A., Buxton R.S., McGrath J.A.;
RT "Genomic organization and amplification of the human desmosomal
RT cadherin genes DSC1 and DSC3, encoding desmocollin types 1 and 3";
RL Biochem. Biophys. Res. Commun. 276:454-460(2000).
CC -!- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.
DR EMBL; AF293339; AAG23427.1; -.
DR HSSP; P15116; INCU.
DR InterPro; IPR002126; Cadherin.
DR Pfam; PF00028; cadherin; 5.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 5.
DR PROSITE; PS00232; CADHERIN_1; 2.
DR PROSITE; PS0268; CADHERIN_2; 5.
DR Calcium-binding; Cell adhesion; Glycoprotein.
KW SEQUENCE 839 AA; 93497 MW; F218E2D8869A6855 CRC64;

Query Match 15.5%; Score 671.5; DB 4; Length 839;
Best Local Similarity 30.8%; Pred. No. 1.3e-35;
Matches 187; Conservative 106; Mismatches 263; Indels 51; Gaps 22;

QY 239 WKAPKPEVMVENSTDPHPKIKITQVRWNPDAQYSL-----VDREKLPRFPFSDQEG 290
Db 137 W-APICSMQENSLGFPFLQOQVE-SDAQNYTVFYSISGRGVDEPLNLFIERD-TG 193
QY 291 DIYVTOPLDREKDAY--VFYAVAKDEYKGLPSYPIEIHVKVDINDNPPTCSPVTVFE 348
Db 194 NLFCTRPVDEEYDVLAYASTADGYSADL--PLPLPIRVEDENDHPVFTEATYNFE 251
QY 349 VOENERLNSIGTLTAHDREENTANSFLNRYVEGTPKLPMDGLFLIQTIYAGNLQAKQ 408
Db 252 VLESSRPGTVGVVCATDRDEPTMTLRKYSILOQTPRSP--GLFSVPSTGVITVSH 309
QY 409 SLKKQDTPYNTLIEVSDRDKF-----TLCFVQINVIDINDQIPIFEKSDYGNLTAE 461
Db 310 YLDREVVDKYSIMKVDMDGQFFGLIGTSCI--ITVTDSDNNAFTFQONAYE--AFVE 365
QY 462 DTNIGSTIITQATDADEPFTGSSKILYHIKGDSEGRIGVDTPHTNTGYVIKKPLDF 521
Db 366 ENAFNVEILRIPTEDKDLINTANWRVFTILKGNENGHEFKISTDKETNEGVLVSVKPLNY 425
QY 522 ETAAVSNIVKAEPELFGV-KYNASFEAKFTLITVDNEAPQFSQHVQAKVSEDVA 580
Db 426 EENRQVNLIEGVNNEAPFARDIPRVTAALRALVTVHVRDLDEGPECTPAAQYVRKENLA 485

```

```

QY 581 IGTKVGNVTAKDPE---GLDISYSLRGDTRGWLKIDHVTGEIFESVAPLDRAGSP----Y 633
Db 486 VGSKINGRAYDENRNGNGLRYKKLHDPKGWITIDEISGSIITSKILDRVEVTPKNELY 545
QY 634 RVQVATEVGGSSLSVSEFHLIMDVNDPRLAKDYTGCLFFCHPLSAPGSLIFEATDD 693
Db 546 NITVLAIDKDRSCTGLAVN--IEDVNDNPPETLOEY--VVICKPKMGVTDIL--AVDP 599
QY 694 DQHLFRGPHFTFSLGSS--LQNDWEVSKINGTHARLS-TRHTDPERAYVVLIRINDGG 750
Db 600 DEPVHAGP-FYFSLPNTSPSISRLSKLVNDTAARLSYQKNAGFOE--YTIPTVKD-- 654
QY 751 RPPLEGIVSLPTFCSCVEGSCFRPAGHOTGIPVGNAGV-GILLTLLVIGIILAVVFI 809
Db 655 RAGQAATKLLRVNLCBCTHTPTQCRATSRSTGVLGKWAAILLIGTALLFSLVLLTLCGV 714
QY 810 IKDKGK 816
Db 715 FGATGK 721

RESULT 7
Q8TBX3
ID Q8TBX3 PRELIMINARY; PRT; 713 AA.
AC Q8TBX3;
DT 01-JUN-2002 (TremBLrel. 21, Created)
DT 01-JUN-2002 (TremBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Cadherin 13, H-cadherin (heart).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC028624; AAH28624.1; -.
SQ SEQUENCE 713 AA; 78320 MW; E947F5FA773988FA CRC64;

Query Match 15.1%; Score 652; DB 4; Length 713;
Best Local Similarity 28.9%; Pred. No. 1.9e-34;
Matches 192; Conservative 100; Mismatches 274; Indels 98; Gaps 21;

QY 160 NGOLYYQIVQLPWINNVYFOINKTKAISL---FREGSOELNPAKNPSYN-----LVI 211
Db 61 NDKLRYEV-----SSPYKVNSDGGLVALRNITAVGKTLFVHARTPHAEDMAELVIV 112
QY 212 SVKDMGQSEN--SFSDTTSV-----DIIVTENIWRAPKPEVMVENSTDPHPKIKITQVRW 264
Db 113 GGRDIOGSLQIDFKFARTSPVPRKRSIVS-----PILPENQRQFPFROVGV-- 162
QY 265 NDPGAQYSLVDKBLPRFPF-----SIDQE-----GDIYVTOPLDREKDAYVF 308
Db 163 -----VDSRPSRKFLTGKVDQEPKGIFRINENTGVSMTRTLDRVIAVYQL 213
QY 309 YAVAKDEYKGLPSYPIEIHVKVDINDNPPTCSPVTVFEVQENERLNSIGTLTAHD 368
Db 214 FVETTVNGKTEGSPLEIVIVIDQNDNRDIFREGPYIGHVMEGSPGTGTTVMRTAFDAD 273
QY 369 EENTANSFLNRYVEOTPKLPMDGLFLIQTIYAG-MLQAKQSLKKQDTP---POYNLTIEV 424
Db 274 DPATDNALLRYNTRQTPDKPSPMFVIDPEKGDIVTVVSPALLDRTLENPKYELIEA 333
QY 425 SDK-----DFKTLCFVQINVIDINDQIPIFEKSDYGNLTAEIDNTGISTILTQATDADE 479
Db 334 QDMAGLDVGLTGATATIMIDDKNDHSPKFTKKEF--QATVEEGAVG-VIVNLTVEDKDD 390
QY 480 PFTGSSKILYHIKGDSEGRIGVDTPHTNTGYVIKKPLDFTAAVSNIVFRAENPEPL 539
Db 391 PATCAWRAAYTIINGNPGQSFEIHTNPQTNEGMLSVVKPLDYETSAPHTLLIKVNEBDPL 450

```

```
Qy 540 VFGVKYNASSPAKFTLIIVTDVNEAPQFSQHVQAKVSEDVAIGTKVGNVTAKDPEGLD-- 597
Db 451 VPDVSYGPSSTATVHITVLDVNEGFVFPDPMMVTRQEDLSVGSVLLTVNATDPSLQHQ 510
Qy 598 -ISYSLRGDTRGWLKIDHVTGEIFSAPLDREA----GSPYRVQVVAPEVSGSSLSVSE 652
Db 511 TIRYSVKDPAGWLNINPINTVDTTAVLDRESTFVDNSVYTFALFLAIDSGNPPATGCTGT 570
Qy 653 FHLILMDVNDPPRLAKDYTGFCFCHPLSAP-----GSLFEATDDOHLFRGPHFT 704
Db 571 LLITLEDVNDAP-----FIYPTVAEVCDDAKNLSVILGASDKDLHPNTDP-FK 619
Qy 705 FSLGSGSLND-WEVSKINGTHARLSTRHTDFFERAYVLLIRINDGRRPPLLEGIVSLPVT 763
Db 620 FEIHKQAVDPKWKISKINNTALVSLLO-NLNKANTNLPIMVTDGSKPPMTNITDLRVQ 678
Qy 764 FCSC 767
Db 679 VCSC 682

RESULT 8
Q8VDK4
ID Q8VDK4 PRELIMINARY; PRT; 714 AA.
AC Q8VDK4;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Cadherin 13.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC021628; AAH21628.1; -.
DR InterPro: IPR002126; Cadherin.
DR Pfam: PF00028; cadherin; 5.
DR PRINTS: PR00205; CADHERIN.
DR SMART: SM00112; CA; 5.
DR PROSITE: PS00232; CADHERIN_1; UNKNOWN_3.
DR PROSITE: PS0268; CADHERIN_2; 3.
SQ SEQUENCE 714 AA; 78116 MW; CASCA791C46A2FEB CRC64;

Query Match 15.1%; Score 652; DB 11; Length 714;
Best Local Similarity 29.3%; Pred. No. 1.9e-34;
Matches 189; Conservative 109; Mismatches 287; Indels 60; Gaps 20;

Qy 160 NGQLYQIVQLPMINNVYFQINNTGATSL---TREGSQELNPAKNPSYN-----LVI 211
Db 61 NEKLHYEV-----SSPHFKVNSDGLVALRNITAVGRTLFVHARTPHAEADMAELVIV 112
Qy 212 SVKDMGQSEN--SFSDTTSV-----DIIVTENIWKAPKPVENVSTDPHPKIQTQVRW 264
Db 113 GGDIGSLQIDIFKFAKTSVPKQKRSIVVS-----PILPENQRPFRDVGKWD 164
Qy 265 ND--PGAQYSLVDK--EKLPFPFSDIQE-GDIYVTPQLDREKDAVYFVAVAKDEYKGP 319
Db 165 SDRPEGSKFLTKGVDQDPKGFRIENNTGVSYSVTFTLDRETIAYLYVETDASGKT 224
Qy 320 LSPLEIHKVVKDINDNPPCPSPVTFVEQENRGNSTGTTTAHREDENTANSFLNY 379
Db 225 LEGVPVLEIVIDQDNRPFRFEGPYIGHVMEGSPGCTTVMRTAFDADPATDNALLRY 284
Qy 380 RIVEQTPKLPMDGLFLIQTIVAG-MLQAKSLAKQDT---POYNLTIEVSDK-----DFK 430
Db 285 NIKQQIPDKPSNFWIDPEKGDIVTVVSPFALLDRETNLENPKYELLIEAQDMAGLDVGLT 344
Qy 431 TLCFVQINVIDINQIPIFEKSDYGNLTAEEDTNIGSTILTITQATDADEPFTGSSKILYH 490
Db 431 TLCFVQINVIDINQIPIFEKSDYGNLTAEEDTNIGSTILTITQATDADEPFTGSSKILYH 490
```

```
Db 345 GTATATIVDDKNDHSPKFTKKEF--QATVEGAVG-VIVNLTVEDKDDPTTGANRAAYT 401
Qy 491 IIKGSEGRGLVDVDPHTNTGYVIKKPLDDETAASVNVFKAENPEPLVFGVKYNASSF 550
Db 402 IINGPGQSFETHNPQTNQEGMLSVWKPLDYEISAFHTLLIKVENEPLVDPDVSYPSSST 461
Qy 551 AKFTLIIVTDVNEAPQFSQHVQAKVSEDVAIGTKVGNVTAKDPEGLD---ISYSLRGDTR 607
Db 462 ATVHITVLDVNEGFVFPDPMMVTRQEDLSVGSVLLTVNATDPSLQHQTIIRYSYKQPA 521
Qy 608 GWLKIDHVTGEIFSAPLDREA----GSPYRVQVVAPEVSGSSLSVSEFHLILMDVNDN 663
Db 522 GWSLNPINGIVDTAVLDRESFPVHNSVYTFALFLAIDSGNPPATGCTGTTLLITLEDINDN 581
Qy 664 PPRIAKDYTGFCFCHPLSAPGSLIFEATDDOHLFRGPHFTFSLGSGSLND-WEVSKIN 722
Db 582 APVIYP--TVAEVCDDARNLSWILGASDKDLHPNTDP-FKFEIHKQAVDPKWKISKIN 638
Qy 723 GTHARLSTRHTDFFERAYVLLIRINDGRRPPLLEGIVSLPVTFCSC 767
Db 639 NTHALVSLLO-NLNKANTNLPIMVTDGSKPPMTNITDLRVQVCS 682

RESULT 9
O77704
ID O77704 PRELIMINARY; PRT; 814 AA.
AC O77704;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Desmocollin type 2 (Fragment).
GN DSC2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Roberts G.A.; Burdett I.D.; Pidsley S.C.; King I.A.; Magee A.I.;
RA Buxton R.S.;
RT "Antisense expression of a desmocollin gene in MDCK cells alters
RT desmosome plaque assembly but does not affect desmoglein expression.";
RL Eur. J. Cell Biol. 76:192-203(1998).
CC -!- SIMILARITY: CONTAINS 4 CADHERIN DOMAINS.
DR EMBL: AJ002299; CAA05309.1; -.
DR HSSP: P09803; 1SUH.
DR InterPro: IPR002126; Cadherin.
DR Pfam: PF00028; cadherin; 5.
DR SMART: SM00112; CA; 4.
DR PROSITE: PS00232; CADHERIN_1; 2.
DR PROSITE: PS0268; CADHERIN_2; 4.
DR KEGG: K04461; Cell adhesion; Glycoprotein.
FT NON_TER 1
FT CHAIN 101 814 DESMOCOLLIN TYPE 2.
SQ SEQUENCE 814 AA; 91081 MW; 59EF086C69BD3882 CRC64;

Query Match 15.1%; Score 651.5; DB 6; Length 814;
Best Local Similarity 30.2%; Pred. No. 2.5e-34;
Matches 181; Conservative 110; Mismatches 256; Indels 53; Gaps 20;

Qy 239 WKAPKPVEMVENVSTDPHPKIQTQVRNNDPGAQYSL-----VDKEKLPFRFPESIDOE- 289
Db 102 W-APIPCSMQENSLGPPFLFLOIQ-SDTAQNTYTFYSIRGPGVDRE--PKNLFYVERDT 157
Qy 290 GDIVVTPQLDREKDAVYFVAVAKDEYKGPLSYPLEIHKVVKDINDNPPCPSPVTFVEV 349
Db 158 GNFUCFPRDREYESFELIAFPATPDGYPPELPPLVIRIEDENDNYPPIFTKTYVFTI 217
Qy 350 QENRGNSTGTTTAHREDENTANSFLNYRIVEQTPKLPMDGLFLIQTIVAGMLQAKQS 409
Db 218 SENCVRGVSTGVCAKTDKDEPDMHTRLKYSIIIEQLPAYPT--LFSMHPATGVTITSSSQ 275
```



Query Match 14.5%; Score 628; DB 11; Length 4589;  
 Best Local Similarity 25.3%; Pred. No. 1.3e-31;  
 Matches 228; Conservative 135; Mismatches 329; Indels 210; Gaps 32;

QY 26 KFSGLPKMTFSIYE---GOEPSQIIFQ-----FKANPPAVTFELTGETDNIFVI 72  
 DB 2703 RFSEPF--YTYTISEDPIGTEIDLIRVHSGTVLYLVKGNTP-----ESNRDEFFVI 2754

QY 73 ERE-GLLYNNRDLRETRSTHNLQVAA---LDANGIIVEGVPVITIEVKDINDNRPTFLQ 128  
 DB 2755 DRQGRLEKSLDHTTKYQFSILARCTLDDEYEVAS--IDVSIQVKDANDNSPVLES 2812

QY 129 SKYEGSVRONSRPKPFLLYNATDLDPPATNGQLYQYIOLPMINNVMYFQINNKTKGA 188  
 DB 2813 NPYEAFIVENLPASRVIOVRASDLDSGV--NGQVMYSL-DQSQDADIIIESFAINMETGW 2869

QY 189 ISLTREGSOELNPAKNPSYNLVISVKDMGQSENSFSDTTSVDIIVTE-----NI 238  
 DB 2870 ITTLK-----ELDHEERASYQIKVVASDHGEKVQ--LSSTAIVGVTVTDVNDSPRFTAEI 2923

QY 239 WKA-----PKPEVMVENSTDPHPKITQVRWNDPGAGYSLVDKELPRFPF-SIDQEG 290  
 DB 2924 YKGTVEDDPPGGVIAILLSTTDADSEINR-----QVSYFITGGDALGQFAVENMQNDW 2977

QY 291 DIYVTOPLDREKDAYFYAVAKDEYKPLSYPLEIHVKVKDINDNPPTC-----340  
 DB 2978 RYVYKKPLDREKQSYLLTVTATD---GTFFSKARVEVKVLDANDNSPVCERTAYSDAIP 3034

QY 341 -----340

DB 3035 EDALPGKLVMOVSATDADIRSNAEITYTLFGSGAEKFKLNPDGTGELRTLALLDREEQAVY 3094

QY 341 -----PSPTVFEVQENRGLNSIGTILTAH 365

DB 3095 HLLVKATDGGRSQCAQVTLTLEDVNDNTPTEFTADPYATVF-----ENTPGTPLTRVQAT 3151

QY 366 DRDEENTANSLNVRIVIEQTPKPLMDGLFIOTVAGMLQAKOSLKKQDTPQYNLTTEVS 425  
 DB 3152 DADA--GLNRKISYSLDSA-----DQFISINEQSGILQLEKH-LDRELOQAVYTLTLKAA 3203

QY 426 D-----KDFKTLCFVQINVIDINDQIPIFERSDYGNLTALAEINTIGSTILTIQATDADPEF 481  
 DB 3204 DOGLPRKLTATGTWVSVLDINDNPNPVFEYREYG-ATVSEDIVIGTEVLQVYAASRD--I 3260

QY 482 TGSSKILYHLIKGSEGLGVDTDPHTNTGVYIKKPLDFETRAVSNIVFKAENPEPLVF 541  
 DB 3261 EANAETIYAIISGHEKFSID-----SKTGAIFIENLDYESSHGYVLTVEATD-----3310

QY 542 GVKNYASSFAKFTLIVTDVNE-APQFSOHVFQAKVSESDVAIGTKVGNVTAKD---PBGLD 597  
 DB 3311 GGTPSLSDVATVINITDINDNPFVSQDSYTTVVSEDALEQPVITIMADDAAGGPNSSH 3370

QY 598 ISYS-LRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVATEVEGGSSLSVSEPHLI 656  
 DB 3371 ILXSIIEGNOGSPFTIDPVGEIKVTKPLDRETISGTVLVQAADNGNPPRVNTITVNID 3430

QY 657 LMDVNDNPPRLANDYGLFFCHPLSAPGLIFEATDDQHLFRGPHFTFSLGSGSLQNDW 716  
 DB 3431 VSDVNDNAPLFSKNDNSVIVIOENKPVGFSVLKLVVTDKDSHNGPPFPFSAIVSGNDDNMF 3490

QY 717 EVSKINGTHARLSTRHTDFE--ERAVVVLIRINDGGRPPLEGIVSLPVTFCVSGSCFR 774  
 DB 3491 EVNQ----HGVLLTAATVRRKVKDHYLLHVKVADNKGPKQLSSLTHIDI----RVIESIHP 3543

QY 775 PA 776  
 DB 3544 PA 3545

RESULT 12

Q9QXA3

ID Q9QXA3

PRELIMINARY; PRT; 4587 AA.

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3

Q9QXA3



```
Db 3096 NLLVKATDGGRCQAIIVLTLEDVNDNAPEFTAEPYITVF---ENTPEGTPLTRVQAT 3152
QY 366 DRDEENTANSELYRIYETPKLPMDGLFIQTYAGMLQALOSLKKQDTPQYNLTIEYS 425
Db 3153 DAD--TGLNRKISYSLVESA-----DQPSINERSGIIQLEKH-LDRELQAVITTLTKAV 3204
QY 426 D----KDFKTLCLFQVINDIQIPIFEKSDYGNLTAEADTNIGSTILTIQATDADEPF 481
Db 3205 DQGLPRLRLTATCTVVVSVLDINDNPPVEYREV-ASVSEDIVIGTEVLQVYAASRD--I 3261
QY 482 TGSSKILYHIIKGDSEGRGLCVDTDHTNGYVLIKKPLDFEFAAANSNIIVFKAENPEPLVF 541
Db 3262 EANAETIYAIISGNEHGKFSID----SKTGAIFIIIESLDYESHSEYLLVEATD----- 3311
QY 542 GYKYNASSFAKTLIVTDVNE-APOFSQHVFOAKYSEDAVIGTKVGNVTAKD---PEGLD 597
Db 3312 GGTPLSDVATVNIINVTINDNSPVFSQDTYTVVSEDAALQOPVITIMADDADGPNSH 3371
QY 598 ISYS-LRGDTRGLWKIDHVTGIFSVAPLDREAGSPYRVQVATEYVGGSSLSVSEFHLI 656
Db 3372 IHYSIEGNOGSPFTIDPVRGEVKVTKPLDRETIISGYTLTVQAADNGNPPRVNTTTVNID 3431
QY 657 LMDVNDNPPRLAKDVTGLFFCHPLSNAPGSLIFEATDDDDHGLRGPHTFSLGSGSLQNDW 716
Db 3432 VSDVNDNAPLFRSDNYSVIIQENKPVGFSLVTLVVDKDSHNGPPFFFTIVSGNDENAF 3491
QY 717 EYKINGTHARLSRHTDPEERA---YVVLIRINDGRRPPLGIVSLPFTFCSCVEGSCF 773
Db 3492 EVNQ---HGVLLTAAAT-IKRKVKDHYFLHVAVDSGKPOLSSMWHIDI---RVIEESH 3543
QY 774 RPA 776
Db 3544 PPA 3546

RESULT 13
Q8R508 PRELIMINARY; PRT; 4555 AA.
AC Q8R508;
DT 01-JUN-2002 (TremBLrel. 21, Created)
DT 01-JUN-2002 (TremBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Fat3.
GN FAT3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus;
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX MEDLINE=21670959; PubMed=11811999;
RA Mitsui K., Nakajima D., Ohara O., Nakayama M.;
RT "Mammalian fat3: a large protein that contains multiple cadherin and
RL EGF-like motifs.";
RL Biochem. Biophys. Res. Commun. 290:1260-1266(2002).
RL EMBL: AB076401; BAB86869.1;
SQ SEQUENCE 4555 AA; 502077 MW; B47CBC10638EA73B CRC64;

Query Match 14.1%; Score 611; DB 11; Length 4555;
Best Local Similarity 25.4%; Pred. NO. 1.7e-30;
Matches 225; Conservative 144; Mismatches 320; Indels 196; Gaps 33;

QY 30 PLKPMTFISYEQEPQSIIFQFKNAPPATVFTLGETDNIFVIERE-GLLYNRLDRET 88
Db 2733 PNQSRFSTVNGERPE-----NNKENVFIEQETGAIKLDRKLDHEV 2774
QY 89 RSTHNLQVAA---LDANGIIVEGPPVITIEVKINDNRPTFLQSKYSGSVNRPGKPF 145
Db 2775 SPAFHFKVAATPLDKVDIVF--TVDDVKVLDLNDNRKPVFTSSYETIIMGPMVGKL 2832
QY 146 LYVNATLDLDDPATPNQGLYQIVIQPLMINVMYFQINNKTKGAISLTRGSOELNPAKP 205
```

```
Db 2833 AQVRAIDTDWGA--NCQVTYSLHSDSHLEKVMFAFNDSNTGWISTLK---DLDETDP 2886
QY 206 SYNLYSVKDMGSGOSENSESDTTSVDIIVTENIWKAPKPVEMVENSTDPHPKIITVRWN 265
Db 2887 TTSFFVVASDLG--EAFSLSSMALVSVKVTINDINAP-----VFAHEVIRGNVKES 2935
QY 266 DRG-----AQSILVDKEKL-----PRPFES-----IDQEGDIYVTOPLDREED 304
Db 2936 DPGGEVAVLSTLKDTSNINRQSVYHITGNPRGRFALGMVQSEKWKVVKRPLDREED 2995
QY 305 AVFVAVAKDEYKPLSYPLEIHVKVDINDNPPTC-----PSPVTVEVQE 351
Db 2996 IYFLNITASD--GLFVTQAM-VEVTVSDVNDNSPVCQVAYSASLPEDIPSNKILKVA 3052
QY 352 NER-----LGNS-----IGTLTAHDRDE----- 369
Db 3053 KDADIGSGDIRSYLSGNSDFLDPESELTKTALLDRERVPVNLNARATDGGGRFC 3112
QY 370 -----ENTANSFLNYRIVEQTKLPMD----- 391
Db 3113 SSTVLLLEDVNDNPPVFSNHYTACVYENTATKALLTRVQAVDPDVGINRKVVYSLEDS 3172
QY 392 --GLFLIQIYAGMLQAKSLKKQDTPQYNLTIEVSD----KDFKTLCLFQVINDIQ 445
Db 3173 ASGVFSDSSGVIVL-EQPLDREQSSYINISVRATDQSPGQSLSLTSVTITVLINDN 3231
QY 446 IPFEKSDYGNLTAEADTNIGSTILTIQATDADEPFTGSSKILYHIKGDSEGRGLVDTD 505
Db 3232 PVFERRDY-LTVVPEDTSLGTQVLSVFATSKD--IGTNAETIYILRSNEQKGFRI--- 3285
QY 506 PHTNTGYVLIKKPLDFETAANSNIVEKAENPEPLVFGVKNYNASFAKTLIVTDVNE-AP 504
Db 3286 -NPKTGGISVLEALDYEMCKRFYLVVEAKD-----GGTPALSTAATVSDLTVDNDNP 3338
QY 565 QFSQHVFOAKYSEDAVIGTKVGNVTAKD---PEGLDISYSL-RGDTGRGLKIDHTVGEI 619
Db 3339 RFSQDYSAVISSEDALEGDVSLIIAEDVDKSPNG-QIRFSIVGGDRDNEFAVDPLGLV 3397
QY 620 FSVAPLDREAGSPYRVQVATEYVGGSSLSVSEFHILMDVNDNPPRLA-KDVTGLFFCH 678
Db 3398 KVKKKLDREVRVSGYSLIIQAVDSGIPAMSTTVTNIDISDVNDNSPVFTPANTAVIQEN 3457
QY 679 PLSAPGLIFEATDDQHLFRGPHFTFSLGSGSLQNDWEVSKINGTHARLSRHTDPEER 738
Db 3458 KPVGTSLQLVVTDRDS-FHNGPPFSFSLSGNEDEEFMLDSHGILRSVAVFRHMSPE- 3515
QY 739 AVVLIRINDGRRPPLGIVSLPFTFCSCVEGSCFRPAGHQTGIP 783
Db 3516 -YLLCTQAKDSGKPKQ---VSHTYIRVRVIEESTHKP-----TAIP 3552

RESULT 14
Q9V5N8 PRELIMINARY; PRT; 3579 AA.
AC Q9V5N8;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE PROTOCADHERIN-like wing polarity protein STAN precursor (STARRY NIGHT
DE protein) (FLAMINGO protein).
GN STAN OR CG11895 OR FMI.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE=EMBRYO;
RX MEDLINE=20025940; PubMed=10556066;
--- RA Chae J.W., Kim M.-J., Goo J.H., Collier S., Gubb D., Charlton J.,
Adler P.N., Park W.J.;
```



Query Match 13.9%; Score 601.5; DB 5; Length 3579;  
 Best Local Similarity 29.1%; Pred. No. 5e-30;  
 Matches 224; Conservative 131; Mismatches 321; Indels 95; Gaps 35;

Qy 37 SIYEGQPSQIFQKAPN-----PAVTELTGTGDN-----IFVIE-REGILY 79  
 Db 472 SIREATVGSVITIRATDQDQIGKNAEYEGIAEVTDCAGLAQDQEMPIFRIDSRSGLV 531  
 Qy 80 YNRALDRETRSHNLOVAALDANGIVE---GPVPITIEVKDINDNRPTFLQSKYEGSVR 136  
 Db 532 TRSSLDRETSQSYHLLTAADLASAQSERRTATASQVQKVLDDNDNYQFSERTVTVQVP 591  
 Qy 137 QNSRPG---KFLYVNAATLDDPATPNQGLYQYQVQLPMNNVMYFQINKTKGALSILTR 193  
 Db 592 EDQWGGTENDTVAHIRATDAQD---GNAAIRYAIIGG---NTQSFSDMSGDSVSLVK 645  
 Qy 194 EGSQELNPAKPSNLYVSKDMGSGENSEFSDTTSVDLIIVTENIWKAPK-----PVEM 247  
 Db 646 ---PLDYESVRSYRLVIRAQDGGSPSR---SNTQLLVNVIDANDNAPRFTTSQFQESV 698  
 Qy 248 VENSTDHPHIKITQVRNDPQAGQ-----YSLVDKEKLPRFPFSID-OEGDIYVTPQLDREE 302  
 Db 699 LENVPVGVNIIRVQAYDSDEGANAETIYSISERO--DNFPLAVDPRTGWVOTIKPLDREE 756  
 Qy 303 KDAYVYAVAKDEYCKPLSYPLEIHVKVDINDNPPTCPSPVTVFEVQENERLGNISGTL 362  
 Db 757 QGRFAQVAVKDGVPVPSASSSVVITVDQNDNDNPAFNPKYEAANGVEDPQGPVPTTV 816  
 Qy 363 TAHDREENTANSFNRYIVTEQTPKLPMDGLFLI--QYVAGMLQAKSLKODPQYNL 420  
 Db 817 TATDPDED---SLHYEITGNTN---GRFAITSQNGRGLITTA-OSLDYKQEKRELL 867  
 Qy 421 TIEVSKDKFT-LCFVQIVINDINDQIPFEKSDYGNLTALBETNIGSTILTIOATDADE 479  
 Db 868 TVAATDSGGRSDTATVHINITDANNFAPFENAPY-SASVEDAPVGTITLVVSATDS- 925  
 Qy 480 PFTGSSKILYHIKDSGRGLGVDTDP---HTNGYVVIKPLDPEETAAGNIVFKAENP 536  
 Db 926 --VGNAQITTSLENSINGLG-SPDPFSINPQTGAIVTNAPLADRETTSGLVLLTVAKD- 981  
 Qy 537 EPLVGVKYNASSFAKFTLIVTDVNE-APQFSQHVQAKVEDVAIGTKVGNVTKADPE- 594  
 Db 982 -----GGNPSLSDTTDVEIGVTDVNDNAPAFKPSPLYQASILEDALVGTSVIQAASDPDV 1036

Qy 595 GLD--ISYSL--RGDTRGWLKIDHVTGEIFSVAPLDRAGSPYRVQVQVATEVGGSSLSV 650  
 Db 1037 GLNGRIKYLSDRDIEDGFSVIDPTSGTIRTKNGKDRSVAVFHLTAITADKRGSPPLST 1096  
 Qy 651 SEFHLLMDVNDNPPRLAKDYTLGFLFCFCHPLSAP-GSLIFE--ATDDOHLFGPHFTFSL 707  
 Db 1097 VEVOIRLEDVNDSPPTFASDKITLV--PENSPVSGVGEIHAHPDDEGVNAVH--YSI 1152  
 Qy 708 GSGSLQNDWE-VSKINGTHARLST-RHTDFE--ERAYVVLIRINDGGRPL 754  
 Db 1153 IGGDDSNASFSLVTRPGSERAQLLTMTELDYESTRKRKFELVVR--AASPL 1200

RESULT 15  
 Q9VW71 PRELIMINARY: PRT: 4643 AA.  
 AC Q9VW71  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Putative fat-like cadherin precursor (CG7749 protein).  
 GN CG7749.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers J., H.C., Blazef R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Besley E.M.,  
 RA Benson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,  
 RA Foster C., Gabrielson A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mettel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Klamas I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E.C., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zhong X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 CC -!- FUNCTION: MAY BE INVOLVED IN CELL ADHESION.  
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).  
 CC -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.  
 CC -!- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.

CC -!- SIMILARITY: CONTAINS 1 LAMININ G-LIKE DOMAIN.  
DR EMBL; AE003515; AAF49078.1; -.  
DR HSSP; P15116; INCI.  
DR FlyBase; FBgn0036930; fat2.  
DR InterPro; IPR000152; ASX\_hydroxyl.  
DR InterPro; IPR002126; Cadherin.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR001881; EGF-CA.  
DR InterPro; IPR001791; Laminin-G.  
DR Pfam; PF00028; cadherin; 31.  
DR Pfam; PF00008; EGF; 5.  
DR Pfam; PF00054; laminin\_G; 1.  
DR PRINTS; PR00205; CADHERIN.  
DR SMART; SM00112; CA; 33.  
DR SMART; SM00179; EGF\_CA; 1.  
DR SMART; SM00001; EGF\_like; 5.  
DR SMART; SM00282; LamG; 1.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 1.  
DR PROSITE; PS00232; CADHERIN\_1; 17.  
DR PROSITE; PS00268; CADHERIN\_2; 41.  
DR PROSITE; PS00022; EGF\_1; 5.  
DR PROSITE; PS01186; EGF\_2; 2.  
DR PROSITE; PS01187; EGF\_CA; 1.  
KW Hypothetical protein; Cell adhesion; Signal; Transmembrane; Repeat;  
KW EGF-like domain; Glycoprotein; Calcium-binding.  
FT SIGNAL 1 35  
FT CHAIN 36 4643  
FT DOMAIN 36 1647 PUTATIVE FAT-LIKE CADHERIN.  
FT TRANSMEM 1648 1668 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 1669 1697 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 168 180 CADHERIN 1.  
FT DOMAIN 189 288 CADHERIN 2.  
FT DOMAIN 293 397 CADHERIN 3.  
FT DOMAIN 406 504 CADHERIN 4.  
FT DOMAIN 513 610 CADHERIN 5.  
FT DOMAIN 619 713 CADHERIN 6.  
FT DOMAIN 778 874 CADHERIN 7.  
FT DOMAIN 883 977 CADHERIN 8.  
FT DOMAIN 986 1085 CADHERIN 9.  
FT DOMAIN 1094 1190 CADHERIN 10.  
FT DOMAIN 1199 1296 CADHERIN 11.  
FT DOMAIN 1305 1402 CADHERIN 12.  
FT DOMAIN 1413 1503 CADHERIN 13.  
FT DOMAIN 1512 1609 CADHERIN 14.  
FT DOMAIN 1618 1714 CADHERIN 15.  
FT DOMAIN 1723 1812 CADHERIN 16.  
FT DOMAIN 1821 1929 CADHERIN 17.  
FT DOMAIN 1948 2030 CADHERIN 18.  
FT DOMAIN 2024 2137 CADHERIN 19.  
FT DOMAIN 2146 2232 CADHERIN 20.  
FT DOMAIN 2247 2338 CADHERIN 21.  
FT DOMAIN 2347 2465 CADHERIN 22.  
FT DOMAIN 2474 2567 CADHERIN 23.  
FT DOMAIN 2576 2670 CADHERIN 24.  
FT DOMAIN 2679 2779 CADHERIN 25.  
FT DOMAIN 2788 2876 CADHERIN 26.  
FT DOMAIN 2885 2983 CADHERIN 27.  
FT DOMAIN 2992 3083 CADHERIN 28.  
FT DOMAIN 3092 3185 CADHERIN 29.  
FT DOMAIN 3194 3289 CADHERIN 30.  
FT DOMAIN 3298 3394 CADHERIN 31.  
FT DOMAIN 3403 3499 CADHERIN 32.  
FT DOMAIN 3508 3604 CADHERIN 33.  
FT DOMAIN 3618 3671 CADHERIN 34.  
FT DOMAIN 3819 3879 EGF-LIKE 1.  
FT DOMAIN 3881 3919 EGF-LIKE 2.  
FT DOMAIN 4129 4166 EGF-LIKE 3.  
FT DOMAIN 4168 4205 EGF-LIKE 4.  
FT DOMAIN 4243 4279 EGF-LIKE 5.  
FT DOMAIN 3823 3835 POTENTIAL.  
FT DISULFID 3830 3867 POTENTIAL.  
FT DISULFID 3869 3878 POTENTIAL.  
FT DISULFID 3885 3896 POTENTIAL.

FT DISULFID 3890 3907 POTENTIAL.  
FT DISULFID 3909 3918 POTENTIAL.  
FT DISULFID 4133 4144 POTENTIAL.  
FT DISULFID 4138 4154 POTENTIAL.  
FT DISULFID 4156 4165 POTENTIAL.  
FT DISULFID 4172 4183 POTENTIAL.  
FT DISULFID 4177 4193 POTENTIAL.  
FT DISULFID 4195 4204 POTENTIAL.  
FT DISULFID 4247 4258 POTENTIAL.  
FT DISULFID 4252 4267 POTENTIAL.  
FT DISULFID 4269 4278 POTENTIAL.  
FT CARBOHYD 65 65 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 156 156 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 364 364 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 779 779 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 843 843 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 923 923 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1106 1106 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1198 1198 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1312 1312 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1439 1439 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1473 1473 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1511 1511 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 4643 AA; 517614 MW; 7EC52BD36ADD4E89 CRC64;  
Query Match 13.7%; Score 591.5; DB 5; Length 4643;  
Best Local Similarity 24.2%; Pred. No. 3.4e-29;  
Matches 210; Conservative 153; Mismatches 312; Indels 193; Gaps 33;  
QY 33 PMTFSTYEQG-----EPSQIIIFQ-FKANPPAVT-----FELTGETDNIFVI 72  
DB 2764 PVSIEIETDANIPTFEKSSVLLKIIESTPGCTVTLKLMIGNYTFKFSIAADQDH-FMI 2822  
QY 73 EREGLLYNRALDRETRSHNLQVAALDANGIIVEGVPITIEVKDINDNRPTFLOSKEYE 132  
DB 2823 SDSGELLQOQLDREQQESHNLIVAAETSTVPVFAYADVLDIVRDENDNPKFEDNTFTS 2882  
QY 133 GSVQRNSRQKPFLLYNATDLDPATPNQOLYYQIVQLPMLNNVMYFOINNKTAISLT 192  
DB 2883 ASVAENSEKVISLVKVSATDAD--TGPNGDIRYILESDTENIQNI--FDIDYSGWITLL 2938  
QY 193 REGSQELNPAKNSYLVISVKDMGQSQNSFSDDTISVDIIVTEN--IWKAP-KPVEMWE 249  
DB 2939 TSLDREV---QSEYNFKVIAADNGHPKHDAKVPVTIKIVDYNDAVPVKLPLEGSLVEE 2994  
QY 250 NSTDPHPKIKITQVRNNDPGAQYSLVDKEKLPRF-----PFSIDQEGDIIVTQPLDRE 301  
DB 2995 NAL-PCGTVLI-NLLLIIDPD-----IEKQEMDFIFVSGDKQAQFQIGKSGELFIAKPLDRE 3047  
QY 302 EKDAYFYAVAKDEYCKPLSYPLEIHVKYKDINDNPCTCPSVTVFEVOENERLNSIGT 361  
DB 3048 QLMFYNLSIATD--GK-FTAKANVEIDYKDINDNTPCYCLKPRYHISTNESISIGTTLVE 3104  
QY 362 LTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQAKQSLKQDTPQYNLT 421  
DB 3105 VRAIDFDQSKLRFYLSGKGAD-----FSGCKESGILKVA-SALDRETPPKYKLV 3154  
QY 422 IEVSD-KDFKTLICF--VQINVIDINDQIPFEKSDYGNLTLAEDTNGSTILTIOATDAD 478  
DB 3155 AHVQDGKDFTECFSEBIIITVNDINDNMPFISMAQY-RVSVPEDAQLNTLITKVHAMDK 3213  
QY 479 EPFTGSSKILYHIKGDSEGRIGVDTPHTNTGYVILKPKLPDETAASVNIKFKE- 534  
DB 3214 ---FGVNRQIKYSLMGENDHYFKIS---KSTGIIRLHKSLODRETISLFLNLTVKAE 3266  
QY 535 -----NPEPLVFGVK----- 544  
DB 3267 PKLHSTATVAVNLLDINDNPPFESMRQYSCCKILENATHGTEVCVKVYATSIDIGVNADIH 3326  
QY 545 -----YNASSF-----AKFTLIVTD 559  
DB 3327 FIMSGNEQGEKMDSTTGDLVLNATLDYEMSKFYFLTIQAIIDGGTPPLSNAYVNISILD 3386



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 2, 2003, 15:46:39 ; Search time 3.31507 Seconds  
(without alignments)  
1768.599 Million cell updates/sec

Title: US-09-079-678-51  
Perfect score: 249  
Sequence: 1 REFAERLWGCDLSWRLDA.....PTPSNRVAKHRKPRSPAL 44

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002.\*  
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	249	100.0	44	AAW74093	Gastro-intestinal
2	249	100.0	44	AAW74069	Human HPT-1 bindin
3	67.5	27.1	486	22 AAG81174	Mycobacterium tube
4	63	25.3	74	22 ABG24044	Novel human diagno
5	63	25.3	88	22 ABG24049	Novel human diagno
6	62.5	25.1	113	22 AAG62344	Propionibacterium
7	61	24.5	116	22 AAG76078	Human colon cancer
8	59.5	23.9	65	22 AAB64712	Human secreted pro
9	59	23.7	252	22 AAU57007	Propionibacterium
10	59	23.7	1093	14 AAR56978	Human myotonic dys

11	58.5	23.5	161	22 AAB98483	Rat RISKMARKER2 re
12	57.5	23.1	387	22 ABG23875	Novel human diagno
13	57.5	23.1	527	22 ABG21299	Novel human diagno
14	57	22.9	61	22 AAG01719	Novel human diagno
15	56.5	22.7	65	22 AAU58690	Propionibacterium
16	56.5	22.7	66	22 AAU66207	Propionibacterium
17	56.5	22.7	106	21 AAG26002	Zea mays protein f
18	56.5	22.7	134	21 AAG26001	Zea mays protein f
19	56.5	22.7	162	21 AAG26000	Zea mays protein f
20	56.5	22.7	305	23 ABP41926	Human ovarian anti
21	56.5	22.7	977	18 AAW36066	Soybean DNA-bindin
22	56	22.5	80	22 ABG27589	Novel human diagno
23	56	22.5	139	22 AAM90960	Human immune/haema
24	56	22.5	155	23 ABP41262	Human ovarian anti
25	56	22.5	1218	22 ABB63445	Drosophila melanog
26	55.5	22.3	63	21 AAG03018	Human secreted pro
27	55.5	22.3	165	22 AAB88604	Human hydrophobic
28	55.5	22.3	165	23 ABB89697	Human polypeptide
29	55.5	22.3	165	23 ABB06127	Human NS protein s
30	55.5	22.3	249	22 AAU32270	Novel human secret
31	55.5	22.3	308	22 AAU51635	Propionibacterium
32	55.5	22.3	356	22 ABB68551	Drosophila melanog
33	55	22.1	59	22 AAU42705	Propionibacterium
34	55	22.1	246	17 AAW05050	Human thrombopoiet
35	55	22.1	282	17 AAW05045	Human thrombopoiet
36	55	22.1	282	17 AAR97701	Human MPL receptor
37	55	22.1	286	16 AAR80827	Human MGF-3. Hom
38	55	22.1	286	17 AAW05044	Human thrombopoiet
39	55	22.1	286	17 AAR91948	Human platelet pro
40	55	22.1	286	17 AAR91964	Megakaryocyte diff
41	55	22.1	286	17 AAR88430	Human MGF-3. Hom
42	55	22.1	286	21 AAY51117	Human platelet-pro
43	55	22.1	286	21 AAY51118	Human platelet-pro
44	55	22.1	401	18 AAW10537	Leptin 1-167/IgG1
45	55	22.1	518	22 AAU28185	Novel human secret

## ALIGNMENTS

RESULT 1  
AAW74093  
ID AAW74093 standard; peptide: 44 AA.  
XX  
AC AAW74093;  
XX  
DT 04-MAY-1999 (first entry)  
XX  
DE Gastro-intestinal transport receptor binding protein.  
XX  
KW Gastro-intestinal transport receptor; binding protein; hSI; HPT1;  
KW D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;  
KW intestinal peptide-associated transporter; hypertension; diabetes;  
KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;  
KW therapeutic agent delivery; therapy.  
XX  
OS Homo sapiens.  
XX  
PN WO9851325-A2.  
XX  
PD 19-NOV-1998.  
XX  
PF 15-MAY-1998; 98WO-US10088.  
XX  
PR 15-MAY-1997; 97US-0046595.  
XX  
PA (CYTO-) CYTOGEN CORP.  
XX  
PA (ELAN-) ELAN CORP PLC.  
PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin IJ;  
PI Omahony DJ, Patterson CA, Singleton J;  
XX  
DR WPI; 1999-009568/01.

XX New proteins that bind specifically to receptors in the  
PT gastro-intestinal tract and related nucleic acid - chimaeras and  
PT antibodies, used to deliver therapeutic or diagnostic agents to, or  
PT through, the gastrointestinal tract, e.g. insulin or leuprolide  
XX  
PS Example; Page 94; 294pp; English.  
XX  
CC The invention relates to purified proteins (I) that bind  
CC specifically to at least one of the gastro-intestinal (GI) tract  
CC receptors human intestinal peptide-associated transporter (HPT1),  
CC hPEPT1, D2H and human sucrose-isomaltase complex (hSI). (I) provide  
CC active transport of therapeutic agents through human and animal GI tissue  
CC (into the blood) for in vivo delivery, particularly for treatment or  
CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,  
CC cancer, migraine, or angina pectoris. Specifically they are used to  
CC deliver insulin or leuprolide, but many other suitable therapeutic agents  
CC are disclosed, including genes or inhibitory nucleic acid, imaging agents  
CC and antigens. (I) may also provide targeting to the GI tract. Other uses  
CC of (I) are: (i) to determine the level of specified receptors in a sample  
CC (in a binding assay); and (ii) to screen for molecules that bind (I).  
CC immunogenic analogues or derivatives of (I) are used to raise antibodies  
CC and in immunoassays. The antibodies are used to locate, detect and  
CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis  
CC etc., also for peptide purification and immobilisation.  
XX  
SQ Sequence 44 AA;  
Query Match 100.0%; Score 249; DB 20; Length 44;  
Best Local Similarity 100.0%; Pred. No. 8.7e-24;  
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 REFERRLLWGCDLDSWRLDAEGCGTPPSNRVAVKHKRPRSPAL 44  
|||||  
Db 1 REFERRLLWGCDLDSWRLDAEGCGTPPSNRVAVKHKRPRSPAL 44  
|||||  
RESULT 2  
AAW74069  
ID AAW74069 standard; peptide; 44 AA.  
XX  
AC AAW74069;  
XX  
DT 04-MAY-1999 (first entry)  
XX  
DE Human HPT-1 binding protein HAX40.  
XX  
KW Gastro-intestinal transport receptor; binding protein; hSI; HPT1;  
KW D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;  
KW intestinal peptide-associated transporter; hypertension; diabetes;  
KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;  
KW therapeutic agent delivery; therapy.  
XX  
OS Homo sapiens.  
XX  
PN WO9851325-A2.  
XX  
PD 19-NOV-1998.  
XX  
PF 15-MAY-1998; 98WO-US10088.  
XX  
PR 15-MAY-1997; 97US-0046595.  
XX  
PA (CYTO-) CYTOGEN CORP.  
PA (ELAN-) ELAN CORP PLC.  
XX  
PI Alvarez VL, Belinka BA, Cagney CM, Carter JM, Lambkin IJ;  
PI Omahony DJ, Patterson CA, Singleton J;  
XX WPI; 1999-009568/01.  
XX  
PT New proteins that bind specifically to receptors in the  
PT gastro-intestinal tract and related nucleic acid - chimaeras and

PT antibodies, used to deliver therapeutic or diagnostic agents to, or  
PT through, the gastrointestinal tract, e.g. insulin or leuprolide  
XX  
PS Claim 2; Page 55; 294pp; English.  
XX  
CC This sequence represents a peptide that specifically binds to the human  
CC HPT-1 protein. The invention relates to purified proteins (I) that bind  
CC specifically to at least one of the gastro-intestinal (GI) tract  
CC receptors human intestinal peptide-associated transporter (HPT1),  
CC hPEPT1, D2H and human sucrose-isomaltase complex (hSI). (I) provide  
CC active transport of therapeutic agents through human and animal GI tissue  
CC (into the blood) for in vivo delivery, particularly for treatment or  
CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,  
CC cancer, migraine, or angina pectoris. Specifically they are used to  
CC deliver insulin or leuprolide, but many other suitable therapeutic agents  
CC are disclosed, including genes or inhibitory nucleic acid, imaging agents  
CC and antigens. (I) may also provide targeting to the GI tract. Other uses  
CC of (I) are: (i) to determine the level of specified receptors in a sample  
CC (in a binding assay); and (ii) to screen for molecules that bind (I).  
CC immunogenic analogues or derivatives of (I) are used to raise antibodies  
CC and in immunoassays. The antibodies are used to locate, detect and  
CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis  
CC etc., also for peptide purification and immobilisation.  
XX  
SQ Sequence 44 AA;  
Query Match 100.0%; Score 249; DB 20; Length 44;  
Best Local Similarity 100.0%; Pred. No. 8.7e-24;  
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 REFERRLLWGCDLDSWRLDAEGCGTPPSNRVAVKHKRPRSPAL 44  
|||||  
Db 1 REFERRLLWGCDLDSWRLDAEGCGTPPSNRVAVKHKRPRSPAL 44  
|||||  
RESULT 3  
AAG81174  
ID AAG81174 standard; Protein; 486 AA.  
XX  
AC AAG81174;  
XX  
DT 04-SEP-2001 (first entry)  
XX  
DE Mycobacterium tuberculosis potential drug target protein SEQ ID 225.  
XX  
KW Drug target; growth; organism viability; characterisation.  
XX  
OS Mycobacterium tuberculosis.  
XX  
PN WO200135317-A1.  
XX  
PD 17-MAY-2001.  
XX  
PF 13-NOV-2000; 2000WO-US31152.  
XX  
PR 12-NOV-1999; 99US-0165086.  
XX  
PR 12-NOV-1999; 99US-0165124.  
XX  
PR 01-FEB-2000; 2000US-0179531.  
XX  
PA (REGC ) UNIV CALIFORNIA.  
XX  
PI Eisenberg D, Rotstein SH, Marcotte EM;  
XX WPI; 2001-329193/34.  
XX N-PSDB; AAH52025.  
XX  
PT Identifying nucleotide or polypeptide sequence for use as drug target,  
PT involves providing algorithm that analyzes a functional relationship  
PT between nucleotide or polypeptide sequences, and comparing the  
PT sequences -  
XX  
PS Disclosure; Page 173-174; 207pp; English.



CC This invention relates to a method for identifying a nucleotide or  
 CC polypeptide sequence that may be a drug target, or essential for growth  
 CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092  
 CC represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium  
 CC tuberculosis proteins which are potential drug targets. The DNA and  
 CC protein sequences are used to illustrate the method of the invention. The  
 CC method involves providing an unknown nucleotide or polypeptide sequences,  
 CC and comparing it to a number of sequences along with at least one  
 CC algorithm capable of analysing a functional relationship between  
 CC nucleotide and polypeptide sequences. The method is useful for  
 CC characterising the function of nucleic acids and polypeptides that may be  
 CC useful as a target for a drug or essential for the growth or viability of  
 CC an organism.

XX Sequence 486 AA;

Query Match 27.1%; Score 67.5; DB 22; Length 486;  
 Best Local Similarity 59.1%; Pred. No. 3.6;  
 Matches 13; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

Qy 8 LMGCDLSWRLDAEGC-GPTPS 28  
 :|| :||||| || || |  
 Db 93 IWGDVELAWRLDAAGCYGPPRS 114

RESULT 4  
 ABG24044  
 ID ABG24044 standard; Protein; 74 AA.

XX AC ABG24044;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #24035.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI: 2001-639362/73.

XX DR N-PSDB; AAS88231.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -

XX PS Claim 20; SEQ ID No 54403; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC disorders involving aberrant protein expression or biological activity.

CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 74 AA;

Query Match 25.3%; Score 63; DB 22; Length 74;  
 Best Local Similarity 38.6%; Pred. No. 1.9;  
 Matches 17; Conservative 6; Mismatches 7; Indels 14; Gaps 2;

Qy 14 LSWRL-DAEGCGTPTSNRAVK-----HRKPRPRSPA 43  
 ||:| : || || : ||  
 Db 16 LSFKLGEVGLPLAALTAGKGLFFLPLESRLHRRPKRSPA 59

RESULT 5  
 ABG24049  
 ID ABG24049 standard; Protein; 88 AA.

XX AC ABG24049;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #24040.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI: 2001-639362/73.

XX DR N-PSDB; AAS88236.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -

XX PS Claim 20; SEQ ID No 54408; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.

polypeptides may be used as antigens in the production of antibodies specific for P. acnes proteins. These antibodies can be used to downregulate expression and activity of P. acnes polypeptides and therefore treat P. acnes infections. The antibodies may also be used as diagnostic agents for determining P. acnes presence, for example, by enzyme linked immunosorbent assay (ELISA).

Note: The sequence data for this patent did not form part of the prior art of the present invention. The sequence data was obtained in electronic format directly from [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).

AA  
SQ Sequence 113 AA;  
Query Match 25.1%; Score 62.5; DB 22; Length 113;  
Best Local Similarity 44.7%; Pred. No. 3.4;  
Matches 17; Conservative : 4; Mismatches 10; Indels 7; Gaps 3;

QY 8 LWGDDLSWRLDAEGCGP--TPSNRAVKHRKP-RPRSP 42  
 ||| ||| | | | : : : ||| ||| |||  
 Db 31 LWG----SWRROHLCVPSPASWRLSGRKPFRPRSP 64

RESULT 7  
AAG76078  
ID AAG76078 standard: Protein: 116 AA:

XX AAG76078:

03-SEP-2001 (first entry)

Human colon cancer antigen protein SEQ ID NO:6842.

Human; colon cancer; colon cancer antigen; diagnosis; detection;  
KW  
AA  
Human; colon cancer; colon cancer antigen; diagnosis; detection;  
KW  
AA  
Human; colorectal carcinoma; chromosome 9.  
KW

AA  
PN  
WO200122920-

05-APR-2001.

28-SEP-2000;

AA  
PR 29-SEP-1999; 99US-0157137.

XX  
PR 03-NOV-66; 0087C9T0-5066PA (HUMA-) HUMAN GENOME SCI INC  
XX

PT RUBEN SM, BARASH SC, BLISE  
XX

DR WPI; 2001-235357/24.  
DR N-PSDB; AAH35483.  
XX

PT Nucleic acids encoding useful for preventing

XX  
PS  
Claim 11: page 82

XX  
CC  
CC  
CC

the proteins are collectively known as collagen antigens have cytostatic activity against cancer antigens

therapy and vaccine production. N and P may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate P expression. For example, N and P may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of P by expressing inactive proteins or to supplement the patients own production of P. Additionally, N may be used to produce the colon cancer-associated Ps, by inserting the nucleic acids into a host cell and culturing the cell to express the proteins. N and P can be used in the prevention, diagnosis and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204 and AAH77789 represent sequences used in the exemplification of the present invention.

CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
 CC missing at time of publication, meaning no sequences are present for  
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.  
 XX  
 XX Sequence 116 AA;

Query Match 24.5%; Score 61; DB 22; Length 116;  
 Best Local Similarity 37.5%; Pred. No. 5.3;  
 Matches 12; Conservative 4; Mismatches 8; Indels 8; Gaps 1;

QY 16 WRLDAGCGGTPPSNR-----VKHRKPRP 39  
 Db 1 WMSCHGLGRTESNRTLLLPWPHLVQHRPKP 32

RESULT 8  
 AAB64712  
 ID AAB64712 standard; Protein; 65 AA.

AC AAB64712;  
 DT 23-MAR-2001 (first entry)

DE Human secreted protein sequence encoded by gene 47 SEQ ID NO:106.

XX Human; secreted protein; diagnosis; cytostatic; antirheumatic;  
 KW antiarthritic; dermatological; cardiant; antiinflammatory; anti-ulcer;  
 KW gastrointestinal; solid tumour; rheumatoid arthritis; psoriasis;  
 KW diabetic retinopathy; myocardial angiogenesis; Crohn's disease;  
 KW ulcer.

OS Homo sapiens.

PN WO200077237-A1.

XX 21-DEC-2000.

PF 01-JUN-2000; 2000WO-US14928.

PR 11-JUN-1999; 99US-0138633.

PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (ROSE/) ROSEN C A.

PI Rosen CA, Ruben SM, Komatsoulis GA;

DR WPI: 2001-071280/08.  
 DR N-PSDB; AAF33083.

XX Nucleic acids encoding 49 human secreted polypeptides, useful for  
 PT preventing, diagnosing and/or treating diseases such as tumors,  
 PT rheumatoid arthritis, psoriasis and diabetic retinopathy -

PS Claim 11; Page 479-480; 520pp; English.

XX The polynucleotide sequences given in AAF33037 to AAF33085 encode the  
 CC human secreted proteins given in AAB64666 to AAB64714. AAB64715 to  
 CC AAB64771 represent human secreted polypeptide sequences and proteins  
 CC homologous to them, which are given in the exemplification of the present  
 CC invention. Human secreted proteins have activities based on the tissues  
 CC and cells the genes are expressed in. Examples of activities include:  
 CC cytostatic; antirheumatic; antiarthritic; dermatological; cardiant;  
 CC antiinflammatory; gastrointestinal; and anti-ulcer. The polynucleotides  
 CC and polypeptides can be used in the prevention, treatment and diagnosis  
 CC of diseases associated with inappropriate polypeptide expression.  
 CC Disorders that may be treated or prevented include solid tumours,  
 CC rheumatoid arthritis, psoriasis, diabetic retinopathy, myocardial  
 CC angiogenesis, Crohn's disease and ulcers. The polynucleotides and their  
 CC complementary sequences may also be used as DNA probes in diagnostic  
 CC assays (e.g. polymerase chain reactions (PCR)) to detect and quantitate  
 CC the presence of similar nucleic acid sequences in samples, and therefore  
 CC which patients may be in need of restorative therapy. The polypeptides  
 CC may also be used as antigens in the production of antibodies against the

CC polypeptide and in assays to identify modulators (agonists and  
 CC antagonists) of polypeptide expression and activity. The anti-polypeptide  
 CC antibodies and antagonists may also be used to down regulate expression  
 CC and activity. AAF33028 to AAF33036 and AAB64665 represent sequences used  
 CC in the exemplification of the present invention.  
 XX

SQ Sequence 65 AA;

Query Match 23.9%; Score 59.5; DB 22; Length 65;  
 Best Local Similarity 31.8%; Pred. No. 4.5;  
 Matches 14; Conservative 2; Mismatches 15; Indels 13; Gaps 1;

QY 14 LSWRLDAEGCGTPPSNRVAVKHKRP-----RPRSPAL 44  
 Db 7 LEWLAAGCGCPSPSARGRRLLWHRSPASEATPRLPTVGA PRGPAL 50

RESULT 9  
 AAU57007  
 ID AAU57007 standard; Protein; 252 AA.

AC AAU57007;

DT 27-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #17903.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.

OS Propionibacterium acnes.

PN WO200181581-A2.

XX 01-NOV-2001.

PF 20-APR-2001; 2001WO-US12865.

PR 21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-208841P.

PR 07-JUL-2000; 2000US-216747P.

XX (CORI-) CORIXA CORP.

XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

DR WPI: 2001-616774/71.

DR N-PSDB; AAS59579.

XX Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -

PS Example 1; SEQ ID NO 18202; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as

CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pot\_sequences.

XX SQ Sequence 252 AA;

Query Match 23.7%; Score 59; DB 22; Length 252;

Best Local Similarity 36.8%; Pred. No. 21;

Matches 14; Conservative 5; Mismatches 11; Indels 8; Gaps 2;

QY 8 LMGCDLWSR---LDAGCGGTP-----SNRAVKHRKP 37

||||| : : : ||||| | : |||

Db 94 LHGDDVVYAGKDVKSAGSGTPRTLKQSTAGFHRSP 131

RESULT 10

AAR56978

ID AAR56978 standard; Protein; 1093 AA.

XX AC AAR56978;

DT 25-FEB-1994 (first entry)

XX Human myotonic dystrophy gene protein.

XX Abnormality; muscular dystrophy; CHR 19; chromosome 19;

KW protein kinase; polymerase chain reaction; brain.

XX Homo sapiens.

FH Key Location/Qualifiers

FT Region 1..1093

FT /note- "encoded by predicted reading frame b,

FT x's in the sequence indicate stop codons

FT in the reading frame"

XX WO9317104-A.

XX 02-SEP-1993.

XX 19-FEB-1993; 93WO-US01545.

XX 20-FEB-1992; 92US-0839255.

XX (MASI ) MASSACHUSETTS INST TECHNOLOGY.

PI Brook JD, Housman DE;

XX WPI; 1993-288410/36.

XX DNA sequence of myotonic dystrophy gene - used to produce probes

PT and identify CHR 19 abnormality and protein kinase responsible

PS Disclosure; Fig 6; 64pp; English.

XX The sequence is that encoded by predicted reading frame b of

CC the human myotonic dystrophy (DM) gene. It may be used in the

CC identification of individuals affected by DM.

XX SQ Sequence 1093 AA;

Query Match 23.7%; Score 59; DB 14; Length 1093;

Best Local Similarity 47.6%; Pred. No. 94;

Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 22 GCGPTPSNRAVKHRKPRSP 42

||||| : : : ||||| | : |||

Db 766 GCGVPAGGARPHAPPPAAP 786

RESULT 11

Novel human diagnostic protein #23866.

AAB98483  
 ID AAB98483 standard; Protein; 161 AA.  
 XX  
 AC AAB98483;  
 XX  
 DT 22-AUG-2001 (first entry)  
 XX  
 XX Rat RISKMARKER2 related protein sequence #1.

Identification; toxic; hepatotoxic; differential gene expression;  
 NSAID; non-steroidal antiinflammatory drug.

OS Rattus norvegicus.

XX WO200138579-A2.

PN 31-MAY-2001.

PD 21-NOV-2000; 2000WO-US32049.

XX 22-NOV-1999; 99US-0166923.

PR 18-FEB-2000; 2000US-0183531.

XX 20-NOV-2000; 2000US-0717321.

XX (CURA-) CURAGEN CORP.

XX Gould-Rothberg BE, Dipippo VA, Ramseh TM, Gerwein RW;

PI WPI; 2001-355948/37.

XX Screening hepatotoxic agent comprises contacting test cell population

PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression

PT with reference population and identifying difference in expression

PT levels -

XX Disclosure; Page 10; 76pp; English.

PS The present invention describes a method of screening a test agent for

CC hepatotoxicity. The method comprises: (a) providing a test cell

CC population comprising a cell capable of expressing one or more nucleic

CC acid sequences selected from the group consisting of RISKMARKER 1-8

CC and INJURYMARKER 1-10; (b) contacting the test cell population with a

CC test agent; (c) measuring expression of one or more of the nucleic

CC acid sequences in the test cell population; (d) comparing the

CC expression of the nucleic acid sequence in the test cell population to

CC the expression of the nucleic acid sequence in an reference cell

CC population comprising at least one cell whose exposure status to a

CC hepatotoxic agent is known; and (e) identifying a difference in

CC expression levels of the RISKMARKER or INJURYMARKER sequences, if

CC present, in the test cell population and reference cell population.

CC The method is useful for identifying a hepatotoxic agent. The present

CC sequence is given in the exemplification of the present invention.

XX SQ Sequence 161 AA;

Query Match 23.5%; Score 58.5; DB 22; Length 161;

Best Local Similarity 63.2%; Pred. No. 15;

Matches 12; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 4 AERRLWGCDDLWSRLDAEG 22

||||| : : : ||||| | : |||

Db 29 ARGALWGCEDIS-RLDAAG 46

RESULT 12

ABG23875

ID ABG23875 standard; Protein; 387 AA.

XX

AC ABG23875;

XX

DT 18-FEB-2002 (first entry)

XX

XX Novel human diagnostic protein #23866.

XX		Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW		food supplement; medical imaging; diagnostic; genetic disorder.
XX		Homo sapiens.
OS		
XX		
PN		WO200175067-A2.
XX		
PD		11-OCT-2001.
XX		
PP		30-MAR-2001; 2001WO-US08631.
XX		
PR		31-MAR-2000; 2000US-0540217.
XX		23-AUG-2000; 2000US-0549167.
XX		(HYSE-) HYSEQ INC.
PA		
PI		Dzmanac RT, Liu C, Tang YT;
XX		
DR		WPI; 2001-639362/73.
XX		N-PSDB; AAS88062.
XX		
PT		New isolated polynucleotide and encoded polypeptides, useful in
PT		diagnostics, forensics, gene mapping, identification of mutations
PT		responsible for genetic disorders or other traits and to assess
PT		biodiversity -
XX		
PS		Claim 20; SEQ ID No 54234; 103pp; English.
XX		
CC		The invention relates to isolated polynucleotide (I) and
CC		polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC		polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC		and gene mapping, and in recombinant production of (II). The
CC		polynucleotides are also used in diagnostics as expressed sequence tags
CC		for identifying expressed genes. (I) is useful in gene therapy techniques
CC		to restore normal activity of (II) or to treat disease states involving
CC		(II). (II) is useful for generating antibodies against it, detecting or
CC		quantitating a polypeptide in tissue, as molecular weight markers and as
CC		a food supplement. (II) and its binding partners are useful in medical
CC		imaging of sites expressing (II). (I) and (II) are useful for treating
CC		disorders involving aberrant protein expression or biological activity.
CC		The polypeptide and polynucleotide sequences have applications in
CC		diagnostics, forensics, gene mapping, identification of mutations
CC		responsible for genetic disorders or other traits to assess biodiversity
CC		and to produce other types of data and products dependent on DNA and
CC		amino acid sequences. ABG00010-ABG30377 represent novel human
CC		diagnostic amino acid sequences of the invention.
CC		Note: The sequence data for this patent did not appear in the printed
CC		specification, but was obtained in electronic format directly from WIPO
CC		at ftp.wipo.int/pub/published_pct_sequences.
XX		
SQ		Sequence 387 AA;
		Query Match 23.1%; Score 57.5; DB 22; Length 387;
		Best Local Similarity 35.4%; Pred. No. 50;
		Matches 17; Conservative 3; Mismatches 17; Indels 11; Gaps 2;
QY		6 RRLMGCDLDSRLDAEGCG-----PTP---SNRAVKHRKRPSP 42
		:   :   :
Db		246 RRPGCEDDAGGLQRDCQQWRVLVLPVPVCGGASPHPKPLDLP 293
RESULT 13		
ABG21299		ID
ID		ABG21299 standard; Protein; 527 AA.
XX		
AC		ABG21299;
XX		
DT		18-FEB-2002 (first entry)
XX		
DE		Novel human diagnostic protein #21290.
XX		
XX		Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW	food supplement; medical imaging; diagnostic; genetic disorder.
XX	
OS	Homio sapiens.
XX	
PN	WO200175067-A2.
XX	
PD	11-OCT-2001.
XX	
XX	30-MAR-2001; 2001WO-US08631.
PF	
XX	
PR	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Drmanac RT, Liu C, Tang YT;
XX	
DR	WFI; 2001-639362/73.
DR	N-PSDB; AAS85486.
XX	
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity -
XX	
XX	Claim 20; SEQ ID NO 51658; 103pp; English.
PS	
XX	
CC	The invention relates to isolated polynucleotide (I) and
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC	and gene mapping, and in recombinant production of (II). The
CC	polynucleotides are also used in diagnostics as expressed sequence tags
CC	for identifying expressed genes. (I) is useful in gene therapy techniques
CC	to restore normal activity of (II) or to treat disease states involving
CC	(II). (II) is useful for generating antibodies against it, detecting or
CC	quantitating a polypeptide in tissue, as molecular weight markers and as
CC	a food supplement. (II) and its binding partners are useful in medical
CC	imaging of sites expressing (II). (I) and (II) are useful for treating
CC	disorders involving aberrant protein expression or biological activity.
CC	The polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. ABG00010-ABG30377 represent novel human
CC	diagnostic amino acid sequences of the invention.
CC	Note: The sequence data for this patent did not appear in the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
SQ	Sequence 527 AA;
	Query Match 23.1%; Score 57.5; DB 22; Length 527;
	Best Local Similarity 42.1%; Pred. No. 68;
	Matches 16; Conservative 5; Mismatches 10; Indels 7; Gaps 2;
QY	7 RLWGCDDLWLRLDAEGCGPTPSNRAVHKRPRRSPAL 44
	: : :         : : :
DB	11 RLWGREKISRK---EGAGPTPQ---ESRRIGARGSAL 41
RESULT 14	
ABG01719	
ID	ABG01719 standard; Protein; 61 AA.
XX	
AC	ABG01719;
XX	
XX	13-FEB-2002 (first entry)
DT	
XX	
DE	Novel human diagnostic protein #1710.
XX	
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder.
XX	

```

OS Homo sapiens.
XX WO200175067-A2.
XX
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540317.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX N-PSDB; AAS65906.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
XX Claim 20; SEQ ID No 32078; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX the polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG0010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 61 AA;
SQ
Query Match 22.9%; Score 57; DB 22; Length 61;
Best Local Similarity 30.2%; Pred. No. 8.6;
Matches 16; Conservative 3; Mismatches 22; Indels 12; Gaps 1;

QY 4 AERRLMG-----CDDLWSRLDAEGCGTPTSNRAVKHRKPRSPAL 44
|:|:| | | | | | | | | | | | | | | |
Db 2 ALRAMAPPALQASAGRRKSSDRSWRRPAGSWRETPTQTSATRSASSRPSGSL 54

RESULT 15
ID AAU58690
XX AAU58690 standard; Protein; 65 AA.
XX
XX AAU58690;
XX
XX 13-FEB-2002 (first entry)
XX
XX Propionibacterium acnes immunogenic protein #19586.
XX
XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX dermatological; osteopathic; neuroprotectant.
XX

```

---

```

OS Propionibacterium acnes.
XX WO200181581-A2.
XX
XX
XX 01-NOV-2001.
XX
XX 20-APR-2001; 2001WO-US12865.
XX
XX 21-APR-2000; 2000US-199047P.
XX
XX 02-JUN-2000; 2000US-208841P.
XX
XX 07-JUL-2000; 2000US-216747P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
XX L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX WPI; 2001-616774/71.
XX N-PSDB; AAS59594.
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
XX vaccinating against and diagnosing infections, especially useful for
XX treating acne vulgaris -
XX
XX Example 1; SEQ ID No 19885; 1069pp; English.
XX
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
XX polypeptides. The proteins and their associated DNA sequences are used in
XX the treatment, prevention and diagnosis of medical conditions caused by
XX P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
XX pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
XX P. acnes is also involved in infections of bone, joints and the central
XX nervous system, however it is particularly involved in the inflammatory
XX lesions associated with acne vulgaris. A method for detecting the
XX presence or absence of P. acnes in a patient comprises contacting a
XX sample with a binding agent that binds to the proteins of the invention
XX and determining the amount of bound protein in the sample. The
XX polypeptides may be used as antigens in the production of antibodies
XX specific for P. acnes proteins. These antibodies can be used to
XX downregulate expression and activity of P. acnes polypeptides and
XX therefore treat P. acnes infections. The antibodies may also be used as
XX diagnostic agents for determining P. acnes presence, for example, by
XX enzyme linked immunosorbent assay (ELISA).
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 65 AA;
SQ
Query Match 22.7%; Score 56.5; DB 22; Length 65;
Best Local Similarity 40.5%; Pred. No. 11;
Matches 15; Conservative 3; Mismatches 18; Indels 1; Gaps 1;

QY 4 AERRLMGCDLWSRLDAEGCGTPTSNRAVKHRKPRPR 40
|:|:| | | | | | | | | | | | | | | |
Db 30 AKRRTHQPDYGSRRLLPGCGPRPSG-LPRHQDPDPR 65

Search completed: April 2, 2003, 15:52:18
Job time : 5.31507 secs

```

GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 2, 2003, 15:51:14 ; Search time 1.85845 Seconds  
(without alignments)  
696.608 Million cell updates/sec

Title: US-09-079-678-51

Perfect score: 249

Sequence: 1 REFERRLWGGDDLSWRLDA.....PTPSRAVKHKRPRRSPAL 44

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pap.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pap.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pap.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pap.\*  
5: /cgn2\_6/ptodata/1/iaa/PTCUS\_COMB.pap.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	55	22.1	265	1	US-08-413-803-27
2	55	22.1	265	1	PCT-US95-03776-29
3	55	22.1	286	1	US-08-321-488A-27
4	55	22.1	1291	4	US-09-150-460B-10
5	55	22.1	1291	4	US-09-220-641-5
6	53	21.3	358	1	US-08-604-913B-11
7	53	21.3	374	1	US-08-095-726-14
8	53	21.3	374	1	US-08-096-623A-14
9	53	21.3	521	1	US-08-276-213-3
10	52	20.9	180	4	US-09-133-341-12
11	51	20.5	100	3	US-09-034-916-17
12	51	20.5	904	4	US-08-632-537-1
13	51	20.5	904	5	PCT-US96-05316-1
14	50.5	20.3	84	4	US-09-177-304-2
15	50	20.1	363	1	US-08-258-261B-5
16	50	20.1	363	1	US-08-456-837-5
17	50	20.1	363	1	US-08-457-342-5
18	50	20.1	363	1	US-08-457-646A-5
19	50	20.1	363	1	US-08-458-076A-5
20	50	20.1	363	1	US-08-457-335A-5
21	50	20.1	363	2	US-08-729-214-5
22	50	20.1	363	3	US-09-028-934-5
23	49	19.7	127	1	US-08-392-828C-39
24	49	19.7	127	3	US-09-330-945-39
25	49	19.7	263	1	US-07-927-071-2
26	49	19.7	491	2	US-08-468-812-8
27	49	19.7	491	4	US-08-590-563-8

28	49	19.7	1063	1	US-08-127-499A-8	Sequence 8, Appli
29	49	19.7	1063	1	US-08-482-847-8	Sequence 8, Appli
30	48.5	19.5	331	3	US-08-793-426A-2	Sequence 2, Appli
31	48.5	19.5	331	4	US-09-294-565-2	Sequence 2, Appli
32	48	19.3	186	4	US-09-475-316A-13	Sequence 13, Appli
33	48	19.3	238	4	US-09-111-470-8	Sequence 8, Appli
34	48	19.3	370	3	US-09-248-335-26	Sequence 26, Appli
35	48	19.3	424	4	US-09-149-476-555	Sequence 555, App
36	48	19.3	616	4	US-09-134-001C-4618	Sequence 4618, Ap
37	48	19.3	992	1	US-08-127-499A-1	Sequence 1, Appli
38	48	19.3	992	1	US-08-482-847-1	Sequence 1, Appli
39	48	19.3	1063	1	US-08-093-453B-3	Sequence 3, Appli
40	47.5	19.1	174	2	US-08-683-262B-60	Sequence 60, Appli
41	47.5	19.1	174	4	US-09-361-707-60	Sequence 2, Appli
42	47.5	19.1	763	1	US-08-473-122-2	Sequence 2, Appli
43	47.5	19.1	763	2	US-08-472-478-2	Sequence 2, Appli
44	47.5	19.1	763	2	US-08-463-081B-8	Sequence 8, Appli
45	47.5	19.1	763	2	US-08-461-379A-8	Sequence 8, Appli

RESULT 1  
US-08-413-803-27  
; Sequence 27, Application US/08413803  
; Patent No. 5766581  
; GENERAL INFORMATION:  
; APPLICANT: Bartley, Timothy D.  
; APPLICANT: Bogenberger, Jakob M.  
; APPLICANT: Bosselman, Robert A.  
; APPLICANT: Hunt, Pamela  
; APPLICANT: Kinstler, Olaf B.  
; APPLICANT: Samal, Babru B.  
; TITLE OF INVENTION: METHODS FOR TREATING MAMMALS WITH  
; TITLE OF INVENTION: MONO-PEGYLATED PROTEINS THAT STIMULATE MEGAKARYOCYTE  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: AMGEN INC.  
; STREET: 1840 DeHavilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: US  
; ZIP: 91320-1789  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/413,803  
; FILING DATE: 30-MAR-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/221,768  
; FILING DATE: 31-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/252,628  
; FILING DATE: 31-MAY-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/321,488  
; FILING DATE: 12-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/347,780  
; FILING DATE: 30-NOV-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cook Ph.D., Robert R.  
; REGISTRATION NUMBER: 31,602  
; REFERENCE/DOCKET NUMBER: A-290D  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 265 amino acids

ALIGNMENTS

1. The first part of the document is a list of names and titles, including "The Hon. Mr. Justice" and "The Hon. Mr. Justice".





Db 230 ERETCCLPITWRTSRLGCP 250

## RESULT 8

US-08-096-623A-14  
; Sequence 14, Application US/08096623A  
; Patent No. 5684238  
; GENERAL INFORMATION:  
; APPLICANT: Ausich, Rodney L.  
; APPLICANT: Brinkhaus, Friedhelm L.  
; APPLICANT: Mukharji, Indrani  
; APPLICANT: Proffitt, John H.  
; APPLICANT: Yarger, James G.  
; APPLICANT: Yen, Huel-Che B.  
; TITLE OF INVENTION: Biosynthesis of Zeaxanthin and  
; TITLE OF INVENTION: Glycosylated Zeaxanthin in Genetically Engineered Hosts  
; NUMBER OF SEQUENCES: 104  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Welsh & Katz, Ltd.  
; STREET: 120 S. Riverside Plaza, 22nd Floor  
; CITY: Chicago  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60606

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/096,623A  
; FILING DATE: 22-JUL-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/805,061  
; FILING DATE: 09-DEC-1991

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/662,921  
; FILING DATE: 28-FEB-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/562,674  
; FILING DATE: 03-AUG-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/525,551  
; FILING DATE: 18-MAY-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/487,613  
; FILING DATE: 02-MAR-1990

ATTORNEY/AGENT INFORMATION:  
; NAME: Gamson, Edward P.  
; REGISTRATION NUMBER: 29,381  
; REFERENCE/DOCKET NUMBER: AMO-006.1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (312) 655-1500

TELEFAX: (312) 655-1501  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 374 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-096-623A-14

Query Match 21.3%; Score 53; DB 1; Length 374;  
Best Local Similarity 42.9%; Pred. No. 29;  
Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 5 ERRWGCDDLWSRLDAEGCGP 25  
||| :||| : |||

Db 230 ERETCCLPITWRTSRLGCP 250

## RESULT 9

US-08-276-213-3  
; Sequence 3, Application US/08276213  
; Patent No. 5536655  
; GENERAL INFORMATION:  
; APPLICANT: Thomas, Steven  
; APPLICANT: Laymon, Robert  
; APPLICANT: Himmel, Michael  
; TITLE OF INVENTION: GENE ENCODING FOR THE E1 ENDOGLUCANASE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: National Renewable Energy Laboratory  
; STREET: 1617 Cole Boulevard  
; CITY: Golden  
; STATE: CO  
; COUNTRY: USA  
; ZIP: 80401-3393

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/276,213  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: O'Connor, Edna  
; REGISTRATION NUMBER: 29,252  
; REFERENCE/DOCKET NUMBER: NREL IR# 94-08  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303)231-1000  
; TELEFAX: (303)231-1098  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 521 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal

US-08-276-213-3

Query Match 21.3%; Score 53; DB 1; Length 521;  
Best Local Similarity 55.6%; Pred. No. 42;  
Matches 10; Conservative 1; Mismatches 5; Indels 2; Gaps 1;

QY 9 WGCDD--LSWRLDAEGCG 24  
||| : ||| : |||

Db 169 WCGDPSIDWRLAERAG 186

## RESULT 10

US-09-133-341-12  
; Sequence 12, Application US/09133341A  
; Patent No. 6162440  
; GENERAL INFORMATION:  
; APPLICANT: Hayward, Diane S.  
; APPLICANT: Ling, Paul D.  
; TITLE OF INVENTION: EBNA2 PEPTIDES AND METHODS OF USING SAME  
; FILE REFERENCE: 87512  
; CURRENT APPLICATION NUMBER: US/09/133,341A  
; CURRENT FILING DATE: 1998-08-13  
; EARLIER APPLICATION NUMBER: WO 97US2243  
; EARLIER FILING DATE: 1997-02-12  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 12  
; LENGTH: 180  
; TYPE: PRT  
; ORGANISM: Unknown

FEATURE:  
OTHER INFORMATION: Description of Unknown Virus Organism: Artificial  
US-09-133-341-12

Query Match 20.9%; Score 52; DB 4; Length 180;  
Best Local Similarity 37.5%; Pred. No. 17;  
Matches 15; Conservative 3; Mismatches 18; Indels 4; Gaps 1;

QY 5 ERLAGCDLSWRLDAEGCGPTPSNRVAVKHKRPRSPAL 44  
| | | | | | | | | | | | | | | | | | | | | |  
Db 73 EHLKAGPDLTPTSFDP----PTPEETVKRVSRRQATL 108

RESULT 11  
US-09-034-916-17  
Sequence 17, Application US/09034916  
Patent No. 6046314  
GENERAL INFORMATION:  
APPLICANT: GEBE, JOHN A.  
APPLICANT: SIADAK, ANTHONY W.  
APPLICANT: ARUFFO, ALEJANDRO A.  
TITLE OF INVENTION: SPALPHA: A NOVEL SCAVENGER RECEPTOR  
TITLE OF INVENTION: CYSTEINE-RICH DOMAIN-CONTAINING POLYPEPTIDE AND MONOCLONAL  
TITLE OF INVENTION: ANTIBODIES THERETO  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ROBINS & ASSOCIATES  
STREET: 90 MIDDLEFIELD ROAD, SUITE 200  
CITY: MENLO PARK  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94025  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/034,916  
FILING DATE: 04-MAR-1998  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/039,956  
FILING DATE: 06-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: ROBINS, ROBERTA L.  
REGISTRATION NUMBER: 33,208  
REFERENCE/DOCKET NUMBER: 5998-0020  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 325-7812  
TELEFAX: (650) 325-7823  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 100 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-034-916-17

Query Match 20.5%; Score 51; DB 3; Length 100;  
Best Local Similarity 50.0%; Pred. No. 12;  
Matches 13; Conservative 0; Mismatches 3; Indels 10; Gaps 3;

QY 9 WG--CDDLWRLDAE-----CCGP 25  
| | | | | | | | | | | | | | | | | | | | | |  
Db 21 WGTVCDD-SWLDLDAQVVCQGLGCGP 45

RESULT 12  
US-08-632-537-1  
Sequence 1, Application US/08632537  
Patent No. 6197497

GENERAL INFORMATION:  
APPLICANT: Goade, Diane E.  
APPLICANT: Bell, Richard  
APPLICANT: Jenison, Steven  
TITLE OF INVENTION: Immunoassay for Herpes Simplex Virus.  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hoffmann, Wasson & Gitler  
STREET: 2361 Jefferson Davis Highway  
CITY: Arlington  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 500 kb storage  
COMPUTER: AOPEN Pentium  
OPERATING SYSTEM: WordPerfect 5.1  
SOFTWARE: WordPerfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/632,537  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/426,604  
FILING DATE: 21-APR-95  
ATTORNEY/AGENT INFORMATION:  
NAME: Buttlmi, Jean A.  
REGISTRATION NUMBER: 24,236  
REFERENCE/DOCKET NUMBER: A5144CIP.SEQ  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)415-0100  
TELEFAX: (703)418-2768  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 904 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
PUBLICATION INFORMATION:  
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 14-110; 295-507; 814-901  
US-08-632-537-1

Query Match 20.5%; Score 51; DB 4; Length 904;  
Best Local Similarity 43.5%; Pred. No. 1.4e+02;  
Matches 10; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 20 AEGCGPTPSNRVAVKHKRPRSP 42  
| | | | | | | | | | | | | | | | | | | | | |  
Db 58 AFGAPPTGDPKPKKPKRPPRP 80

RESULT 13  
PCT-US96-05316-1  
Sequence 1, Application PC/TUS9605316  
GENERAL INFORMATION:  
APPLICANT: Goade, Diane E.  
APPLICANT: Bell, Richard  
APPLICANT: Jenison, Steven  
TITLE OF INVENTION: Immunoassay for Herpes Simplex Virus.  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hoffmann, Wasson & Gitler  
STREET: 2361 Jefferson Davis Highway  
CITY: Arlington  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 500 kb storage  
COMPUTER: AOPEN Pentium  
OPERATING SYSTEM: WordPerfect 5.1  
SOFTWARE: WordPerfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/05316



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 2, 2003, 15:54:14 ; Search time 2.26027 Seconds  
(without alignments)  
1190.115 Million cell updates/sec

Title: US-09-079-678-51

Perfect score: 249

Sequence: 1 REFAERLWGCDLRLDA.....PTPSRAVKHKRPRSPAL 44

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 248812 seqs, 61136040 residues

Total number of hits satisfying chosen parameters: 248812

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

#### Database :

Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	67.5	27.1	486	9	US-09-712-363-225
2	56.5	22.7	399	10	US-09-907-372-1
3	56.5	22.7	435	10	US-09-907-372-19
4	56.5	22.7	435	10	US-09-768-779A-6
5	55	22.1	401	10	US-09-859-361-9
6	55	22.1	738	10	US-09-978-979-6
7	54.5	21.9	1056	9	US-10-161-510-10
8	54	21.7	101	9	US-09-820-843A-92
9	53	21.3	562	10	US-09-981-900B-5
10	52.5	21.1	248	9	US-09-925-299-958
11	52.5	21.1	248	10	US-09-925-299-958
12	52.5	21.1	416	9	US-10-124-429-2
13	52	20.9	180	10	US-09-739-852-12
14	52	20.9	380	9	US-09-955-999-82
15	51.5	20.7	240	10	US-09-765-213A-10
16	51.5	20.7	341	9	US-09-971-490-15
17	51.5	20.7	375	9	US-09-738-626-6565
18	51	20.5	448	9	US-09-738-626-4036
19	51	20.5	1116	9	US-09-977-577-10

Sequence 11, Appl  
Sequence 13, Appl  
Sequence 12, Appl  
Sequence 21, Appl  
Sequence 1336, Ap  
Sequence 4, Appl  
Sequence 812, App  
Sequence 214, App  
Sequence 234, App  
Sequence 2, Appl  
Sequence 42064, A  
Sequence 1674, App  
Sequence 294, App  
Sequence 8, Appl  
Sequence 262, App  
Sequence 32, Appl  
Sequence 32, Appl  
Sequence 32, Appl  
Sequence 32, Appl  
Sequence 6726, Ap  
Sequence 1208, Ap  
Sequence 164, App  
Sequence 254, App  
Sequence 75, Appl  
Sequence 255, App

1149 9 US-09-977-577-11  
1151 9 US-09-977-577-13  
1156 9 US-09-977-577-12  
3298 9 US-10-149-819-21  
84 10 US-09-925-300-1336  
360 10 US-09-943-718-4  
72 9 US-10-102-806-812  
204 10 US-09-893-737-214  
204 10 US-09-893-737-234  
204 10 US-09-809-567-2  
738 10 US-09-864-761-42064  
57 10 US-09-867-550-1674  
150 10 US-09-912-020-294  
206 10 US-09-770-621-8  
491 10 US-09-893-737-262  
124 10 US-10-227-884-32  
194 9 US-10-230-163-32  
194 9 US-10-218-631-32  
194 9 US-10-230-338-32  
194 9 US-10-230-414-32  
217 9 US-09-738-626-6726  
288 10 US-09-925-300-1208  
743 10 US-09-771-161A-164  
743 10 US-09-771-161A-254  
912 9 US-10-047-542-75  
1036 10 US-09-771-161A-255

#### ALIGNMENTS

RESULT 1  
US-09-712-363-225  
; Sequence 225, Application US/09712363  
; Patent No. US20020164588A1  
; GENERAL INFORMATION:  
; APPLICANT: Eisenberg, David  
; APPLICANT: Rotstein, Sergio H.  
; APPLICANT: Marcotte, Edward M.  
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND  
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS  
; FILE REFERENCE: 07419-032001  
; CURRENT APPLICATION NUMBER: US/09/712,363  
; CURRENT FILING DATE: 2000-11-13  
; PRIOR APPLICATION NUMBER: PCT/US00/02246  
; PRIOR FILING DATE: 2000-01-28  
; PRIOR APPLICATION NUMBER: 60/179,531  
; PRIOR FILING DATE: 2000-02-01  
; PRIOR APPLICATION NUMBER: 60/117,844  
; PRIOR FILING DATE: 1999-01-29  
; PRIOR APPLICATION NUMBER: 60/118,206,  
; PRIOR FILING DATE: 1999-02-01  
; PRIOR APPLICATION NUMBER: 60/126,593  
; PRIOR FILING DATE: 1999-03-26  
; PRIOR APPLICATION NUMBER: 60/134,093  
; PRIOR FILING DATE: 1999-05-14  
; PRIOR APPLICATION NUMBER: 60/134,092  
; PRIOR FILING DATE: 1999-05-14  
; PRIOR APPLICATION NUMBER: 60/165,124  
; PRIOR FILING DATE: 1999-11-12  
; PRIOR APPLICATION NUMBER: 60/165,086  
; PRIOR FILING DATE: 1999-11-12  
; NUMBER OF SEQ ID NOS: 292  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 225  
; LENGTH: 486  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-09-712-363-225

Query Match 27.1%; Score 67.5; DB 9; Length 486;  
Best Local Similarity 59.1%; Pred. No. 1.4;  
Matches 13; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

QY 8 LMGCDLWSRLDAEGC-GPTPS 28  
:||:||||| || ||  
Db 93 IWGDVELAWRLDAAGCVGPPRS 114

RESULT 2  
US-09-907-372-1  
; Sequence 1, Application US/09907372  
; Patent No. US20020068242A1  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti G.  
; TITLE OF INVENTION: TNF RECEPTOR 2 RELATED PROTEIN VARIANT  
; FILE REFERENCE: PC-0050 US  
; CURRENT APPLICATION NUMBER: US/09/907,372  
; CURRENT FILING DATE: 2001-07-27  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PERL Program  
; SEQ ID NO 1  
; LENGTH: 399  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No. US20020068242A1 7497867CD1  
US-09-907-372-1

Query Match 22.7%; Score 56.5; DB 10; Length 399;  
Best Local Similarity 44.0%; Pred. No. 25;  
Matches 11; Conservative 5; Mismatches 8; Indels 1; Gaps 1;

QY 12 DDLWSRL-DAEGCGTPSNRAVKHR 35  
:||:||||| || ||  
Db 370 DGRAWHLAETEHCGATPSNRGPRNQ 394

RESULT 3  
US-09-907-372-19  
; Sequence 19, Application US/09907372  
; Patent No. US20020068242A1  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti G.  
; TITLE OF INVENTION: TNF RECEPTOR 2 RELATED PROTEIN VARIANT  
; FILE REFERENCE: PC-0050 US  
; CURRENT APPLICATION NUMBER: US/09/907,372  
; CURRENT FILING DATE: 2001-07-27  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PERL Program  
; SEQ ID NO 19  
; LENGTH: 435  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No. US20020068242A1 g339762  
US-09-907-372-19

Query Match 22.7%; Score 56.5; DB 10; Length 435;  
Best Local Similarity 44.0%; Pred. No. 28;  
Matches 11; Conservative 5; Mismatches 8; Indels 1; Gaps 1;

QY 12 DDLWSRL-DAEGCGTPSNRAVKHR 35  
:||:||||| || ||  
Db 406 DGRAWHLAETEHCGATPSNRGPRNQ 430

RESULT 4  
US-09-768-779A-6  
; Sequence 6, Application US/09768779A  
; Patent No. US20020127637A1  
; GENERAL INFORMATION:

APPLICANT: NI, JIAN  
MOORE, PAUL  
TITLE OF INVENTION: HUMAN TUMOR NECROSIS FACTOR  
RECEPTOR-LIKE PROTEIN 8  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HUMAN GENOME SCIENCES, INC.  
STREET: 9410 KEY WEST AVENUE  
CITY: ROCKVILLE  
STATE: MD  
COUNTRY: US  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/09/768,779A  
APPLICATION NUMBER: US/09/768,779A  
FILING DATE: 25-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/086,582  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: KENLEY K. HOOVER  
REGISTRATION NUMBER: 40,302  
REFERENCE/DOCKET NUMBER: PF368PP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8439  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 435 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 6:

US-09-768-779A-6

Query Match 22.7%; Score 56.5; DB 10; Length 435;  
Best Local Similarity 44.0%; Pred. No. 28;  
Matches 11; Conservative 5; Mismatches 8; Indels 1; Gaps 1;

QY 12 DDLWSRL-DAEGCGTPSNRAVKHR 35  
:||:||||| || ||  
Db 406 DGRAWHLAETEHCGATPSNRGPRNQ 430

RESULT 5  
US-09-859-361-9  
; Sequence 9, Application US/09859361  
; Patent No. US20020058311A1  
; GENERAL INFORMATION:  
; APPLICANT: Browne, Michael  
Chapman, Conrad  
Clinkenbeard, Helen  
Robinson, Jeffrey  
TITLE OF INVENTION: Chimeric Leptin Fused to Immunoglobulin  
Domain and Use  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS



US-09-925-299-958



Db 140 EOREKLSYCGVGTW-----VNSGYPYTNRLAFASFDEKNKYKNDLKNTSPRP 186

RESULT 13

US-09-739-852-12  
; Sequence 12, Application US/09739852  
; Patent No. US20010014337A1

GENERAL INFORMATION:

APPLICANT: Hayward, Diane S.  
APPLICANT: Ling, Paul D.  
TITLE OF INVENTION: EBNA2 PEPTIDES AND METHODS OF USING SAME  
FILE REFERENCE: 87512  
CURRENT APPLICATION NUMBER: US/09/739,852  
CURRENT FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: 09/133,341  
PRIOR FILING DATE: 2000-12-18  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 12  
LENGTH: 180  
TYPE: PRT  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: Description of Unknown Virus Organism: Artificial

US-09-739-852-12

Query Match 20.9%; Score 52; DB 10; Length 180;  
Best Local Similarity 37.5%; Pred. No. 39;  
Matches 15; Conservative 3; Mismatches 18; Indels 4; Gaps 1;

QY 5 ERLWCCDLSWRDAGCGTPPSNRAVKRKRPRRSPAL 44

Db 73 EHLKAGDLLTSPFDP-----PTPEETVRKRVSRPQATL 108

RESULT 14

US-09-955-999-82  
; Sequence 82, Application US/09955999  
; Publication No. US20030036505A1

GENERAL INFORMATION:

APPLICANT: Barash et al.  
TITLE OF INVENTION: Signal Transduction Pathway Component Polynucleotides, Polypeptides  
FILE OF INVENTION: Antibodies, and Methods Based Thereon  
FILE REFERENCE: PT086P1  
CURRENT APPLICATION NUMBER: US/09/955,999  
CURRENT FILING DATE: 2001-09-20  
PRIOR APPLICATION NUMBER: 60/234,997  
PRIOR FILING DATE: 2000-09-25  
NUMBER OF SEQ ID NOS: 139  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 82  
LENGTH: 380  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (118)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (132)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (365)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-955-999-82

Query Match 20.9%; Score 52; DB 9; Length 380;  
Best Local Similarity 37.5%; Pred. No. 85;  
Matches 12; Conservative 2; Mismatches 18; Indels 0; Gaps 0;

QY 13 DLSWRDAGCGCTPSNRAVKRKRPRRSPAL 44

Db 263 DDTWLQSASSHSPTQRRPVSSTHPPRPAPV 294

RESULT 15

US-09-765-213A-10  
; Sequence 10, Application US/09765213A  
; Patent No. US20020079846A1

GENERAL INFORMATION:

APPLICANT: Facchini, Peter J  
TITLE OF INVENTION: Polypeptides and Methods of Use  
FILE REFERENCE: 22542-001  
CURRENT APPLICATION NUMBER: US/09/765,213A  
CURRENT FILING DATE: 2001-01-17  
PRIOR APPLICATION NUMBER: 60/176708  
PRIOR FILING DATE: 2000-01-18  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 10  
LENGTH: 240  
TYPE: PRT  
ORGANISM: Zea mays  
US-09-765-213A-10

Query Match 20.7%; Score 51.5; DB 10; Length 240;  
Best Local Similarity 45.5%; Pred. No. 61;  
Matches 15; Conservative 0; Mismatches 15; Indels 3; Gaps 2;

QY 2 EFAERRLWGC--DDLW-RLDAGCGCTPSNRA 31

Db 100 ERAMHRFTAFIDDKFWPALDAVSLAPTCARA 132

Search completed: April 2, 2003, 16:05:36  
Job time : 4.26027 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 2, 2003, 15:52:30 ; Search time 23.105 seconds  
(without alignments)  
1227.796 Million cell updates/sec

Title: US-09-079-678-51  
Perfect score: 249  
Sequence: 1 REFERRLWGCDDLWSRLDA.....PTPSNRAVKHRRPRSPAL 44

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues  
Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main:\*

1: /cgn2\_6/ptodata/1/paa/PCTUS\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/paa/US06\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/paa/US07\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/paa/US080\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/paa/US081\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/paa/US082\_COMB.pep.\*  
7: /cgn2\_6/ptodata/1/paa/US083\_COMB.pep.\*  
8: /cgn2\_6/ptodata/1/paa/US084\_COMB.pep.\*  
9: /cgn2\_6/ptodata/1/paa/US085\_COMB.pep.\*  
10: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep.\*  
11: /cgn2\_6/ptodata/1/paa/US087\_COMB.pep.\*  
12: /cgn2\_6/ptodata/1/paa/US088\_COMB.pep.\*  
13: /cgn2\_6/ptodata/1/paa/US089\_COMB.pep.\*  
14: /cgn2\_6/ptodata/1/paa/US090\_COMB.pep.\*  
15: /cgn2\_6/ptodata/1/paa/US091\_COMB.pep.\*  
16: /cgn2\_6/ptodata/1/paa/US092\_COMB.pep.\*  
17: /cgn2\_6/ptodata/1/paa/US093\_COMB.pep.\*  
18: /cgn2\_6/ptodata/1/paa/US094\_COMB.pep.\*  
19: /cgn2\_6/ptodata/1/paa/US095\_COMB.pep.\*  
20: /cgn2\_6/ptodata/1/paa/US096\_COMB.pep.\*  
21: /cgn2\_6/ptodata/1/paa/US097\_COMB.pep.\*  
22: /cgn2\_6/ptodata/1/paa/US098\_COMB.pep.\*  
23: /cgn2\_6/ptodata/1/paa/US099\_COMB.pep.\*  
24: /cgn2\_6/ptodata/1/paa/US100\_COMB.pep.\*  
25: /cgn2\_6/ptodata/1/paa/US101\_COMB.pep.\*  
26: /cgn2\_6/ptodata/1/paa/US102\_COMB.pep.\*  
27: /cgn2\_6/ptodata/1/paa/US60\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				Query		Description	
Result No.	Score	Match	Length	DB	ID		
1	249	100.0	44	1	PCT-US98-10088-51	Sequence 51, Appl	
2	249	100.0	44	1	PCT-US98-10088-168	Sequence 168, App	
3	249	100.0	44	14	US-09-079-678-51	Sequence 51, Appl	
4	249	100.0	44	14	US-09-079-678-168	Sequence 168, App	
5	249	100.0	44	14	US-09-079-678-51	Sequence 51, Appl	
6	249	100.0	44	14	US-09-079-678-168	Sequence 168, App	

7	249	100.0	44	14	US-09-079-723-51	Sequence 51, Appl
8	249	100.0	44	14	US-09-079-723-168	Sequence 168, App
9	249	100.0	44	14	US-09-079-723-51	Sequence 51, Appl
10	249	100.0	44	14	US-09-079-723-168	Sequence 168, App
11	249	100.0	44	14	US-09-079-819-51	Sequence 51, Appl
12	249	100.0	44	14	US-09-079-819-168	Sequence 168, App
13	249	100.0	44	14	US-09-079-819-51	Sequence 51, Appl
14	249	100.0	44	14	US-09-079-819-168	Sequence 168, App
15	249	100.0	44	23	US-09-443-780C-67	Sequence 67, Appl
16	249	100.0	45	14	US-09-079-678-272	Sequence 272, App
17	249	100.0	45	14	US-09-079-723-272	Sequence 272, App
18	249	100.0	45	14	US-09-079-819-272	Sequence 272, App
19	67.5	27.1	486	21	US-09-712-363-225	Sequence 225, App
20	66	26.5	123	20	US-09-620-111B-791	Sequence 791, App
21	65	26.1	109	21	US-09-708-427-77965	Sequence 77965, A
22	65	26.1	148	21	US-09-708-427-44119	Sequence 44119, A
23	64	25.7	81	21	US-09-708-427-44260	Sequence 44260, A
24	64	25.7	98	21	US-09-708-427-44259	Sequence 44259, A
25	63	25.3	74	1	PCT-US01-08631-54403	Sequence 54403, A
26	63	25.3	88	1	PCT-US01-08631-54408	Sequence 54408, A
27	62	24.9	1131	27	US-60-167-217-20384	Sequence 20384, A
28	61.5	24.7	452	21	US-09-791-537-47615	Sequence 47615, A
29	61	24.5	116	1	PCT-US00-26524B-6842	Sequence 6842, Ap
30	61	24.5	116	25	US-10-106-698-6852	Sequence 6852, Ap
31	60.5	24.3	145	20	US-09-689-980-2507	Sequence 2507, Ap
32	60.5	24.3	152	20	US-09-617-682A-10442	Sequence 10442, A
33	60.5	24.3	160	20	US-09-689-980-2506	Sequence 2506, Ap
34	60.5	24.3	177	21	US-09-708-427-71834	Sequence 71834, A
35	60.5	24.3	349	23	US-09-902-540-10790	Sequence 10790, A
36	60	24.1	189	18	US-09-478-081-1016	Sequence 1016, Ap
37	59.5	23.9	64	23	US-09-950-083-4444	Sequence 4444, Ap
38	59.5	23.9	64	25	US-10-105-299-5448	Sequence 5448, Ap
39	59.5	23.9	65	1	PCT-US00-14928-106	Sequence 106, App
40	59.5	23.9	125	20	US-09-688-052-2303	Sequence 2303, Ap
41	59.5	23.9	141	20	US-09-688-052-2302	Sequence 2302, Ap
42	59	23.7	727	23	US-09-902-540-12383	Sequence 12383, A
43	59	23.7	1107	4	US-08-023-612A-15	Sequence 15, Appl
44	59	23.7	1225	16	US-09-252-991A-25018	Sequence 25018, A
45	58.5	23.5	161	21	US-09-717-321A-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1  
PCT-US98-10088-51  
; Sequence 51, Application PC/TUS9810088  
; GENERAL INFORMATION:  
; APPLICANT: CYTOGEN CORPORATION  
; APPLICANT: ELAN CORPORATION, PLC  
; TITLE OF INVENTION: RANDOM PEPTIDES THAT BIND TO GASTRO-  
; TITLE OF INVENTION: INTESTINAL TRACT (GIT) TRANSPORT RECEPTORS AND RELATED METH  
; NUMBER OF SEQUENCES: 265  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION NUMBER: PCT/US98/10088  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Misrock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 1101-209

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 44 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
PCT-US98-10088-51

Query Match 100.0%; Score 249; DB 1; Length 44;

Best Local Similarity 100.0%; Pred. No. 1.2e-22;

Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 REFAERRLWGCDDLWRLDAEGCGPTPSNRAVKHKKRPRSPAL 44

DB 1 REFAERRLWGCDDLWRLDAEGCGPTPSNRAVKHKKRPRSPAL 44

#### RESULT 2

PCT-US98-10088-168

Sequence 168, Application PC/TUS9810088

GENERAL INFORMATION:

APPLICANT: CYTOGEN CORPORATION

APPLICANT: ELAN CORPORATION, PLC

TITLE OF INVENTION: RANDOM PEPTIDES THAT BIND TO GASTRO-

TITLE OF INVENTION: INTESTINAL TRACT (GIT) TRANSPORT RECEPTORS AND RELATED METHODS

NUMBER OF SEQUENCES: 265

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSEQ Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US98/10088

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Mistrock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 1101-209

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-790-9090

TELEFAX: 212-869-9741

INFORMATION FOR SEQ ID NO: 168:

SEQUENCE CHARACTERISTICS:

LENGTH: 44 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: unknown

MOLECULE TYPE: peptide

PCT-US98-10088-168

Query Match 100.0%; Score 249; DB 1; Length 44;

Best Local Similarity 100.0%; Pred. No. 1.2e-22;

Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 REFAERRLWGCDDLWRLDAEGCGPTPSNRAVKHKKRPRSPAL 44

DB 1 REFAERRLWGCDDLWRLDAEGCGPTPSNRAVKHKKRPRSPAL 44

#### RESULT 3

US-09-079-678-51

Sequence 51, Application US/09079678

GENERAL INFORMATION:

APPLICANT: Alvarez, Vernon L.

APPLICANT: O'Mahony, Daniel J.

APPLICANT: Lambkin, Imelda J.

APPLICANT: Singleton, Judith

APPLICANT: Patterson, Catherine A.

APPLICANT: Cagney, Gerard M.

APPLICANT: Belinka, Benjamin A.

APPLICANT: Carter, John M.

TITLE OF INVENTION: RANDOM PEPTIDES THAT BIND TO GASTRO-

TITLE OF INVENTION: INTESTINAL TRACT (GIT) TRANSPORT RECEPTORS AND RELATED METH

NUMBER OF SEQUENCES: 265

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSEQ Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/079,678

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Mistrock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 1101-220

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-790-9090

TELEFAX: 212-869-9741

INFORMATION FOR SEQ ID NO: 51:

SEQUENCE CHARACTERISTICS:

LENGTH: 44 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-09-079-678-51

Query Match 100.0%; Score 249; DB 14; Length 44;

Best Local Similarity 100.0%; Pred. No. 1.2e-22;

Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 REFAERRLWGCDDLWRLDAEGCGPTPSNRAVKHKKRPRSPAL 44

DB 1 REFAERRLWGCDDLWRLDAEGCGPTPSNRAVKHKKRPRSPAL 44

#### RESULT 4

US-09-079-678-168

Sequence 168, Application US/09079678

GENERAL INFORMATION:

APPLICANT: Alvarez, Vernon L.

APPLICANT: O'Mahony, Daniel J.

APPLICANT: Lambkin, Imelda J.

APPLICANT: Singleton, Judith

APPLICANT: Patterson, Catherine A.

APPLICANT: Cagney, Gerard M.

APPLICANT: Belinka, Benjamin A.

APPLICANT: Carter, John M.

TITLE OF INVENTION: RANDOM PEPTIDES THAT BIND TO GASTRO-

TITLE OF INVENTION: INTESTINAL TRACT (GIT) TRANSPORT RECEPTORS AND RELATED METH

NUMBER OF SEQUENCES: 265

CORRESPONDENCE ADDRESS:

ADDRESS: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
FILING DATE: 15-May-1998  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Misrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-220  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-9741  
INFORMATION FOR SEQ ID NO: 168:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 44 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: peptide  
MOLECULE TYPE: peptide  
US-09-079-678-168

Query Match 100.0%; Score 249; DB 14; Length 44;  
Best Local Similarity 100.0%; Pred. No. 1.2e-22;  
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 REFAERRLWGCDLSWRLDAEGCGTPPSNRVAVKHKRPRSPAL 44  
|||||  
Db 1 REFAERRLWGCDLSWRLDAEGCGTPPSNRVAVKHKRPRSPAL 44

## RESULT 5

US-09-079-678-51  
Sequence 51, Application US/09079678A  
GENERAL INFORMATION:  
APPLICANT: Alvarez, Vernon L.  
O'Mahony, Daniel J.  
Lambkin, Imelda J.  
Singleton, Judith  
Patterson, Catherine A.  
Cagney, Gerard M.  
Belinka, Benjamin A.  
Carter, John M.

TITLE OF INVENTION: RANDOM PEPTIDES THAT BIND TO GASTRO-

NUMBER OF SEQUENCES: 407

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/079,678A  
FILING DATE: 15-May-1998  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:

NAME: Misrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-220  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-9741  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 44 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 51:  
US-09-079-678-51

Query Match 100.0%; Score 249; DB 14; Length 44;  
Best Local Similarity 100.0%; Pred. No. 1.2e-22;  
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 REFAERRLWGCDLSWRLDAEGCGTPPSNRVAVKHKRPRSPAL 44  
|||||  
Db 1 REFAERRLWGCDLSWRLDAEGCGTPPSNRVAVKHKRPRSPAL 44

## RESULT 6

US-09-079-678-168  
Sequence 168, Application US/09079678A  
GENERAL INFORMATION:  
APPLICANT: Alvarez, Vernon L.  
O'Mahony, Daniel J.  
Lambkin, Imelda J.  
Singleton, Judith  
Patterson, Catherine A.  
Cagney, Gerard M.  
Belinka, Benjamin A.  
Carter, John M.

TITLE OF INVENTION: RANDOM PEPTIDES THAT BIND TO GASTRO-

NUMBER OF SEQUENCES: 407

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/079,678A  
FILING DATE: 15-May-1998  
CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Misrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-220  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-9741  
TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 168:

SEQUENCE CHARACTERISTICS:

LENGTH: 44 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: unknown

MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 168:

US-09-079-678-168

Query Match 100.0%; Score 249; DB 14; Length 44;  
Best Local Similarity 100.0%; Pred. No. 1.2e-22;  
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 REFAERRLWGCDLWSRLDAEGCGPTPSNRVAVKHKRPRRSPAL 44  
|||||  
Db 1 REFAERRLWGCDLWSRLDAEGCGPTPSNRVAVKHKRPRRSPAL 44  
|||||

## RESULT 7

US-09-079-723-51  
; Sequence 51, Application US/09079723  
; GENERAL INFORMATION:  
; APPLICANT: Alvarez, Vernon L.  
; APPLICANT: O'Mahony, Daniel J.  
; APPLICANT: Lambkin, Imelda J.  
; APPLICANT: Singleton, Judith  
; APPLICANT: Patterson, Catherine A.  
; APPLICANT: Cagney, Gerard M.  
; APPLICANT: Belinka, Benjamin A.  
; APPLICANT: Carter, John M.  
; TITLE OF INVENTION: RANDOM PEPTIDES THAT BIND TO GASTRO-  
; TITLE OF INVENTION: INTESTINAL TRACT (GIT) TRANSPORT RECEPTORS AND RELATED METHODS  
; NUMBER OF SEQUENCES: 265  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/079,723  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Misrock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 1101-219  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 51:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 44 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-09-079-723-51

Query Match 100.0%; Score 249; DB 14; Length 44;  
Best Local Similarity 100.0%; Pred. No. 1.2e-22;  
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 REFAERRLWGCDLWSRLDAEGCGPTPSNRVAVKHKRPRRSPAL 44  
|||||  
Db 1 REFAERRLWGCDLWSRLDAEGCGPTPSNRVAVKHKRPRRSPAL 44  
|||||

## RESULT 8

US-09-079-723-168  
; Sequence 168, Application US/09079723  
; GENERAL INFORMATION:  
; APPLICANT: Alvarez, Vernon L.

; APPLICANT: O'Mahony, Daniel J.  
; APPLICANT: Lambkin, Imelda J.  
; APPLICANT: Singleton, Judith  
; APPLICANT: Patterson, Catherine A.  
; APPLICANT: Cagney, Gerard M.  
; APPLICANT: Belinka, Benjamin A.  
; APPLICANT: Carter, John M.  
; TITLE OF INVENTION: RANDOM PEPTIDES THAT BIND TO GASTRO-  
; TITLE OF INVENTION: INTESTINAL TRACT (GIT) TRANSPORT RECEPTORS AND RELATED METH  
; NUMBER OF SEQUENCES: 265  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/079,723  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Misrock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 1101-219  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 168:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 44 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-09-079-723-168

Query Match 100.0%; Score 249; DB 14; Length 44;  
Best Local Similarity 100.0%; Pred. No. 1.2e-22;  
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 REFAERRLWGCDLWSRLDAEGCGPTPSNRVAVKHKRPRRSPAL 44  
|||||  
Db 1 REFAERRLWGCDLWSRLDAEGCGPTPSNRVAVKHKRPRRSPAL 44  
|||||

## RESULT 9

US-09-079-723-51  
; Sequence 51, Application US/09079723A  
; GENERAL INFORMATION:  
; APPLICANT: Alvarez, Vernon L.  
; APPLICANT: O'Mahony, Daniel J.  
; APPLICANT: Lambkin, Imelda J.  
; APPLICANT: Singleton, Judith  
; APPLICANT: Patterson, Catherine A.  
; APPLICANT: Cagney, Gerard M.  
; APPLICANT: Belinka, Benjamin A.  
; APPLICANT: Carter, John M.  
; TITLE OF INVENTION: RANDOM PEPTIDES THAT BIND TO GASTRO-  
; TITLE OF INVENTION: INTESTINAL TRACT (GIT) TRANSPORT RECEPTORS AND RELATED M  
; NUMBER OF SEQUENCES: 407  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA

```
;
;
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/079,723A
; FILING DATE: 15-May-1998
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-219
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-9741
;
; TELEFAX: 212-869-9741
;
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 51:
US-09-079-723-51

Query Match 100.0%; Score 249; DB 14; Length 44;
Best Local Similarity 100.0%; Pred. No. 1.2e-22;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 REFAERLWGCDLWSRLDAEGCGTPPSNRVAVKHKRPRSPAL 44
DB 1 REFAERLWGCDLWSRLDAEGCGTPPSNRVAVKHKRPRSPAL 44

RESULT 10
US-09-079-723-168
; Sequence 168, Application US/09079723A
; GENERAL INFORMATION:
; APPLICANT: Alvarez, Vernon L.
; O'Mahony, Daniel J.
; Lambkin, Imelda J.
; Singleton, Judith
; Patterson, Catherine A.
; Cagney, Gerard M.
; Belinka, Benjamin A.
; Carter, John M.
;
; TITLE OF INVENTION: RANDOM PEPTIDES THAT BIND TO GASTRO-
; INTESTINAL TRACT (GIT) TRANSPORT RECEPTORS AND RELATED METH
;
; NUMBER OF SEQUENCES: 407
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/079,723A
; FILING DATE: 15-May-1998
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-219
; TELECOMMUNICATION INFORMATION:
```

```
;
;
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-9741
;
; TELEFAX: 212-869-9741
;
; INFORMATION FOR SEQ ID NO: 168:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 168:
US-09-079-723-168

Query Match 100.0%; Score 249; DB 14; Length 44;
Best Local Similarity 100.0%; Pred. No. 1.2e-22;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 REFAERLWGCDLWSRLDAEGCGTPPSNRVAVKHKRPRSPAL 44
DB 1 REFAERLWGCDLWSRLDAEGCGTPPSNRVAVKHKRPRSPAL 44

RESULT 11
US-09-079-819-51
; Sequence 51, Application US/09079819
; GENERAL INFORMATION:
; APPLICANT: Alvarez, Vernon L.
; O'Mahony, Daniel J.
; Lambkin, Imelda J.
; Singleton, Judith
; Patterson, Catherine A.
; Cagney, Gerard M.
; Belinka, Benjamin A.
; Carter, John M.
;
; TITLE OF INVENTION: RANDOM PEPTIDES THAT BIND TO GASTRO-
; INTESTINAL TRACT (GIT) TRANSPORT RECEPTORS AND RELATED METH
;
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/079,819
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-209
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-9741
;
; TELEFAX: 212-869-9741
;
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 51:
US-09-079-819-51

Query Match 100.0%; Score 249; DB 14; Length 44;
Best Local Similarity 100.0%; Pred. No. 1.2e-22;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 REFAERLWGCDLWSRLDAEGCGPTPSNRAVKHRKPRPRSPAL 44  
|||||  
Db 1 REFAERLWGCDLWSRLDAEGCGPTPSNRAVKHRKPRPRSPAL 44

## RESULT 12

US-09-079-819-168  
; Sequence 168, Application US/09079819  
; GENERAL INFORMATION:  
; APPLICANT: Alvarez, Vernon L.  
; APPLICANT: O'Mahony, Daniel J.  
; APPLICANT: Lambkin, Imelda J.  
; APPLICANT: Singleton, Judith  
; APPLICANT: Patterson, Catherine A.  
; APPLICANT: Cagney, Gerard M.  
; APPLICANT: Belinka, Benjamin A.  
; APPLICANT: Carter, John M.

; TITLE OF INVENTION: RANDOM PEPTIDES THAT BIND TO GASTRO-  
; INTESINAL TRACT (GIT) TRANSPORT RECEPTORS AND RELATED METHODS  
; NUMBER OF SEQUENCES: 265

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/079,819  
; FILING DATE:

; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Misrock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 1101-209  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-9741

; INFORMATION FOR SEQ ID NO: 168:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 44 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-09-079-819-168

Query Match 100.0%; Score 249; DB 14; Length 44;  
Best Local Similarity 100.0%; Pred. No. 1.2e-22;  
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 REFAERLWGCDLWSRLDAEGCGPTPSNRAVKHRKPRPRSPAL 44  
|||||  
Db 1 REFAERLWGCDLWSRLDAEGCGPTPSNRAVKHRKPRPRSPAL 44

## RESULT 13

US-09-079-819-51  
; Sequence 51, Application US/09079819A  
; GENERAL INFORMATION:

; APPLICANT: Alvarez, Vernon L.  
; APPLICANT: O'Mahony, Daniel J.  
; APPLICANT: Lambkin, Imelda J.  
; APPLICANT: Singleton, Judith  
; APPLICANT: Patterson, Catherine A.  
; APPLICANT: Cagney, Gerard M.

; Belinka, Benjamin A.  
; Carter, John M.  
; TITLE OF INVENTION: RANDOM PEPTIDES THAT BIND TO GASTRO-  
; INTESINAL TRACT (GIT) TRANSPORT RECEPTORS AND RELATED M  
; NUMBER OF SEQUENCES: 407  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/079,819A  
; FILING DATE: 15-May-1998

; CLASSIFICATION: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Misrock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 1101-209

; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-9741  
; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 51:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 44 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: <Unknown>  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
; SEQUENCE DESCRIPTION: SEQ ID NO: 51:  
US-09-079-819-51

Query Match 100.0%; Score 249; DB 14; Length 44;  
Best Local Similarity 100.0%; Pred. No. 1.2e-22;  
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 REFAERLWGCDLWSRLDAEGCGPTPSNRAVKHRKPRPRSPAL 44  
|||||  
Db 1 REFAERLWGCDLWSRLDAEGCGPTPSNRAVKHRKPRPRSPAL 44

## RESULT 14

US-09-079-819-168  
; Sequence 168, Application US/09079819A  
; GENERAL INFORMATION:

; APPLICANT: Alvarez, Vernon L.  
; APPLICANT: O'Mahony, Daniel J.  
; APPLICANT: Lambkin, Imelda J.  
; APPLICANT: Singleton, Judith  
; APPLICANT: Patterson, Catherine A.  
; APPLICANT: Cagney, Gerard M.  
; APPLICANT: Belinka, Benjamin A.  
; APPLICANT: Carter, John M.

; TITLE OF INVENTION: RANDOM PEPTIDES THAT BIND TO GASTRO-  
; INTESINAL TRACT (GIT) TRANSPORT RECEPTORS AND RELATED M

## NUMBER OF SEQUENCES: 407

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036

## COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible



```

; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/079,819A
; FILING DATE: 15-May-1998
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-209
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 168:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 168:
US-09-079-819-168

```

```

Query Match      100.0%; Score 249; DB 14; Length 44;
Best Local Similarity 100.0%; Pred. No. 1.2e-22;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 REFAERRLWGCDDLWRDLAEGCGTPPSNRAVKHRKPRSPAL 44
   ||||||||||||||||||||||||||||||||||||||||
DB 1 REFAERRLWGCDDLWRDLAEGCGTPPSNRAVKHRKPRSPAL 44

```

```

RESULT 15
US-09-443-780C-67
; Sequence 67, Application US/09443780C
; GENERAL INFORMATION:
; APPLICANT: O'Mahony, Daniel J
; APPLICANT: Seveso, Michela
; TITLE OF INVENTION: Antibodies To Peptides That Target GIT Receptors And Related Meth
; FILE REFERENCE: E1067/20037
; CURRENT APPLICATION NUMBER: US/09/443,780C
; CURRENT FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: US 60/109,036
; PRIOR FILING DATE: 1998-11-19
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 67
; LENGTH: 44
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: HAX40 44 mer fragment L-form
US-09-443-780C-67

```

```

Query Match      100.0%; Score 249; DB 23; Length 44;
Best Local Similarity 100.0%; Pred. No. 1.2e-22;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 REFAERRLWGCDDLWRDLAEGCGTPPSNRAVKHRKPRSPAL 44
   ||||||||||||||||||||||||||||||||||||||||
DB 1 REFAERRLWGCDDLWRDLAEGCGTPPSNRAVKHRKPRSPAL 44

```

```

Search completed: April 2, 2003, 16:03:01
Job time : 25.105 secs

```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 2, 2003, 15:52:54 ; Search time 4.82192 Seconds  
(without alignments)  
1318.206 Million cell updates/sec

Title: US-09-079-678-51

Perfect score: 249

Sequence: 1 REFERRLWGCDLWRLDA.....PTPSNRAVKHRKPRRSPAL 44

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 656186 seqs, 144460914 residues

Total number of hits satisfying chosen parameters: 656186

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending\_Patents\_AA\_New.\*

1: /cgn2\_6/ptodata/2/paa/PCT\_NEW\_COMB.pep.\*

2: /cgn2\_6/ptodata/2/paa/US06\_NEW\_COMB.pep.\*

3: /cgn2\_6/ptodata/2/paa/US07\_NEW\_COMB.pep.\*

4: /cgn2\_6/ptodata/2/paa/US08\_NEW\_COMB.pep.\*

5: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep.\*

6: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep.\*

7: /cgn2\_6/ptodata/2/paa/US60\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	249	100.0	44	5	US-09-443-986C-66
2	67.5	27.1	486	6	US-10-282-122A-64653
3	67.5	27.1	493	6	US-10-282-122A-62328
4	62.5	25.1	113	1	PCT-US02-32727-23539
5	62.5	25.1	113	5	US-09-978-825-23539
6	62.5	25.1	113	6	US-10-057-498-23539
7	59.5	23.9	585	7	US-60-452-680-13166
8	59	23.7	228	5	US-09-724-676-70801
9	59	23.7	228	5	US-09-724-676A-70801
10	59	23.7	252	1	PCT-US02-32727-18202
11	59	23.7	252	5	US-09-978-825-18202
12	59	23.7	252	6	US-10-057-498-18202
13	59	23.7	555	5	US-09-724-676-70781
14	59	23.7	555	5	US-09-724-676-70782
15	59	23.7	555	5	US-09-724-676A-70781
16	59	23.7	555	5	US-09-724-676A-70782
17	59	23.7	565	5	US-09-724-676-70799
18	59	23.7	565	5	US-09-724-676-70803
19	59	23.7	565	5	US-09-724-676A-70799
20	59	23.7	565	5	US-09-724-676A-70803
21	59	23.7	590	5	US-09-724-676-70793
22	59	23.7	590	5	US-09-724-676-70794
23	59	23.7	590	5	US-09-724-676A-70793
24	59	23.7	590	5	US-09-724-676A-70794
25	59	23.7	595	5	US-09-724-676-70783
26	59	23.7	595	5	US-09-724-676A-70783

ALIGNMENTS

RESULT 1

US-09-443-986C-66

; Sequence 66, Application US/09443986C

; GENERAL INFORMATION:

; APPLICANT: Elan Corporation

; APPLICANT: O'Mahony, Daniel J.

; TITLE OF INVENTION: RETRO-INVERSION PEPTIDES THAT TARGET GIT TRANSPORT RECEPTORS A

; FILE OF INVENTION: METHODS

; FILE REFERENCE: 25,478-A USA

; CURRENT APPLICATION NUMBER: US/09/443,986C

; CURRENT FILING DATE: 1999-11-19

; NUMBER OF SEQ ID NOS: 85

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 66

; LENGTH: 44

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: HAX40 44 mer fragment L-form

US-09-443-986C-66

Query Match

Best Local Similarity 100.0%; Score 249; DB 5; Length 44;

Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 REFERRLWGCDLWRLDAEGCGPTPSNRAVKHRKPRRSPAL 44

Db 1 REFERRLWGCDLWRLDAEGCGPTPSNRAVKHRKPRRSPAL 44

RESULT 2

US-10-282-122A-64653

; Sequence 64653, Application US/10282122A

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

```
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-03-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 64653
; LENGTH: 486
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-282-122A-64653

Query Match 27.1%; Score 67.5; DB 6; Length 486;
Best Local Similarity 59.1%; Pred. No. 30;
Matches 13; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

QY 8 LMGCDLWSRLDAEGC-GPTPS 28
Db 93 IMGDVELAWRLDAAGCYGPPRS 114

RESULT 3
; Sequence 62328, Application US/10282122A
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELIFRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
```

```
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62328
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Mycobacterium bovis
US-10-282-122A-62328

Query Match 27.1%; Score 67.5; DB 6; Length 493;
Best Local Similarity 59.1%; Pred. No. 30;
Matches 13; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

QY 8 LMGCDLWSRLDAEGC-GPTPS 28
Db 100 IMGDVELAWRLDAAGCYGPPRS 121

RESULT 4
; Sequence 32727-23539
; Sequence 23539, Application PC/TUS0232727
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer
; APPLICANT: Skeiky, Yasir
; APPLICANT: Persing, David
; APPLICANT: Bhatia, Ajay
; APPLICANT: Maisonneuve, Jean Francois
; APPLICANT: Zhang, Yanni
; APPLICANT: Wang, Siqing
; APPLICANT: Jen, Shyian
; APPLICANT: Lodes, Michael
; APPLICANT: Benson, Darin
; APPLICANT: Jones, Robert
; APPLICANT: Carter, Darrick
; APPLICANT: Barth, Brenda
; APPLICANT: Douglass, John
; TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acne
; FILE REFERENCE: 210121.514C1
; CURRENT APPLICATION NUMBER: PCT/US02/32727
; CURRENT FILING DATE: 2002-10-11
; NUMBER OF SEQ ID NOS: 30992
; SEQ ID NO 23539
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Propioni acnes
PCT-US02-32727-23539

Query Match 25.1%; Score 62.5; DB 1; Length 113;
Best Local Similarity 44.7%; Pred. No. 33;
Matches 17; Conservative 4; Mismatches 10; Indels 7; Gaps 3;

QY 8 LMGCDLWSRLDAEGCGP--TPSNRAVKHRKP-RPRSP 42
Db 31 LWG----SWRRQHLCVPSVSPASWRLSGRKKPFRPRSP 64

RESULT 5
; Sequence 978-825-23539
; Sequence 23539, Application US/09978825
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer
; APPLICANT: Skeiky, Yasir
; APPLICANT: Persing, David
; APPLICANT: Bhatia, Ajay
; APPLICANT: Maisonneuve, Jean Francois
; APPLICANT: Zhang, Yanni
; APPLICANT: Wang, Siqing
```

```

; APPLICANT: Jen, Shyian
; APPLICANT: Lodes, Michael
; APPLICANT: Benson, Darin
; APPLICANT: Jones, Robert
; APPLICANT: Carter, Darrick
; APPLICANT: Barth, Brenda
; APPLICANT: Douglass, John
; TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acne
; FILE REFERENCE: 210121.514C1
; CURRENT APPLICATION NUMBER: US/09/978,825
; CURRENT FILING DATE: 2003-01-29
; NUMBER OF SEQ ID NOS: 30992
; SEQ ID NO 23539
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Propioni acnes
US-09-978-825-23539

Query Match      25.1%; Score 62.5; DB 5; Length 113;
Best Local Similarity 44.7%; Pred. No. 33;
Matches 17; Conservative 4; Mismatches 10; Indels 7; Gaps 3;

QY 8 LAGCDDLRLDAEGCGP--TPSNRAVKHKP-RPRSP 42
Db 31 LMG-----SWRRQHLLCVSPSPASWRLSGRKPRPRSP 64

RESULT 6
US-10-057-498-23539
; Sequence 23539, Application US/10057498
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer
; APPLICANT: Skeiky, Yasir
; APPLICANT: Persing, David
; TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acne
; FILE REFERENCE: 210121.514
; CURRENT APPLICATION NUMBER: US/10/057,498
; CURRENT FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 29212
; SEQ ID NO 23539
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Propioni acnes
US-10-057-498-23539

Query Match      25.1%; Score 62.5; DB 6; Length 113;
Best Local Similarity 44.7%; Pred. No. 33;
Matches 17; Conservative 4; Mismatches 10; Indels 7; Gaps 3;

QY 8 LMGDDLRLDAEGCGP--TPSNRAVKHKP-RPRSP 42
Db 31 LMG-----SWRRQHLLCVSPSPASWRLSGRKPRPRSP 64

RESULT 7
US-60-452-680-13166
; Sequence 13166, Application US/60452680
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: GRUPE, Andrew
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001450
; CURRENT APPLICATION NUMBER: US/60/452,680
; CURRENT FILING DATE: 2003-03-07
; NUMBER OF SEQ ID NOS: 116213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13166
; LENGTH: 585
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-452-680-13166
```

```

Query Match      23.9%; Score 59.5; DB 7; Length 585;
Best Local Similarity 41.0%; Pred. No. 2.3e+02;
Matches 16; Conservative 1; Mismatches 11; Indels 11; Gaps 2;

QY 4 AERRLWGCDLRLDAEGCGPTPSNRAVKHKRPRPRSP 42
Db 4 AERR-----WROQRGSSPHSLRST---APRPRPP 31

RESULT 8
US-09-724-676-70801
; Sequence 70801, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 70801
; LENGTH: 228
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-70801

Query Match      23.7%; Score 59; DB 5; Length 228;
Best Local Similarity 47.6%; Pred. No. 1.3e+02;
Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 22 GCGPTPSNRAVKHKRPRPRSP 42
Db 164 GCGPVPAGGARPHAPPPPAAP 184

RESULT 9
US-09-724-676A-70801
; Sequence 70801, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 70801
; LENGTH: 228
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-70801

Query Match      23.7%; Score 59; DB 5; Length 228;
Best Local Similarity 47.6%; Pred. No. 1.3e+02;
Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 22 GCGPTPSNRAVKHKRPRPRSP 42
Db 164 GCGPVPAGGARPHAPPPPAAP 184

RESULT 10
PCT-US02-32727-18202
; Sequence 18202, Application PC/TUS0232727
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer
; APPLICANT: Skeiky, Yasir
; APPLICANT: Persing, David
; APPLICANT: Bhatia, Ajay
; APPLICANT: Maisonneuve, Jean Francois
; APPLICANT: Zhang, Fanni
; APPLICANT: Wang, Siqing
; APPLICANT: Jen, Shyian
```

```
; APPLICANT: Lodes, Michael
; APPLICANT: Benson, Darin
; APPLICANT: Jones, Robert
; APPLICANT: Carter, Darrick
; APPLICANT: Barth, Brenda
; APPLICANT: Douglass, John
; TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acne
; FILE REFERENCE: 210121.514C1
; CURRENT APPLICATION NUMBER: PCT/US02/32727
; CURRENT FILING DATE: 2002-10-11
; NUMBER OF SEQ ID NOS: 30992
; SEQ ID NO 18202
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Propioni acnes
PCT-US02-32727-18202

Query Match      23.7%; Score 59; DB 1; Length 252;
Best Local Similarity 36.8%; Pred. No. 1.4e+02;
Matches 14; Conservative 5; Mismatches 11; Indels 8; Gaps 2;

QY 8 LMGCCDLSWR---LDAEGCGPTP-----SNRAVKHRKP 37
| | | | : : : | | | | | : | | |
Db 94 LHGCCDDVVYAGKDVKSAGSGPTPTRTLKQSTAGFRHRSP 131

RESULT 11
US-09-978-825-18202
; Sequence 18202, Application US/09978825
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer
; APPLICANT: Skeiky, Yasir
; APPLICANT: Persing, David
; APPLICANT: Bhatia, Ajay
; APPLICANT: Maisonneuve, Jean Francois
; APPLICANT: Zhang, Yanni
; APPLICANT: Wang, Siqing
; APPLICANT: Lodes, Michael
; APPLICANT: Benson, Darin
; APPLICANT: Jones, Robert
; APPLICANT: Carter, Darrick
; APPLICANT: Barth, Brenda
; APPLICANT: Douglass, John
; TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acne
; FILE REFERENCE: 210121.514C1
; CURRENT APPLICATION NUMBER: US/09/978,825
; CURRENT FILING DATE: 2003-01-29
; NUMBER OF SEQ ID NOS: 30992
; SEQ ID NO 18202
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Propioni acnes
US-09-978-825-18202

Query Match      23.7%; Score 59; DB 5; Length 252;
Best Local Similarity 36.8%; Pred. No. 1.4e+02;
Matches 14; Conservative 5; Mismatches 11; Indels 8; Gaps 2;

QY 8 LMGCCDLSWR---LDAEGCGPTP-----SNRAVKHRKP 37
| | | | : : : | | | | | : | | |
Db 94 LHGCCDDVVYAGKDVKSAGSGPTPTRTLKQSTAGFRHRSP 131

RESULT 12
US-10-057-498-18202
; Sequence 18202, Application US/10057498
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer
; APPLICANT: Skeiky, Yasir
; APPLICANT: Persing, David
; TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acne
; FILE REFERENCE: 210121.514
```

```
; CURRENT APPLICATION NUMBER: US/10/057,498
; CURRENT FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 29212
; SEQ ID NO 18202
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Propioni acnes
US-10-057-498-18202

Query Match      23.7%; Score 59; DB 6; Length 252;
Best Local Similarity 36.8%; Pred. No. 1.4e+02;
Matches 14; Conservative 5; Mismatches 11; Indels 8; Gaps 2;

QY 8 LMGCCDLSWR---LDAEGCGPTP-----SNRAVKHRKP 37
| | | | : : : | | | | | : | | |
Db 94 LHGCCDDVVYAGKDVKSAGSGPTPTRTLKQSTAGFRHRSP 131

RESULT 13
US-09-724-676-70781
; Sequence 70781, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 70781
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-70781

Query Match      23.7%; Score 59; DB 5; Length 555;
Best Local Similarity 47.6%; Pred. No. 2.5e+02;
Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 22 GCGPTPSNRAVKHRKPRRSP 42
| | | | : : : | | | | | : | | |
Db 531 GCGPVPAGGARPHAPPPAAP 551

RESULT 14
US-09-724-676-70782
; Sequence 70782, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 70782
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-70782

Query Match      23.7%; Score 59; DB 5; Length 555;
Best Local Similarity 47.6%; Pred. No. 2.5e+02;
Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 22 GCGPTPSNRAVKHRKPRRSP 42
| | | | : : : | | | | | : | | |
Db 531 GCGPVPAGGARPHAPPPAAP 551

RESULT 15
US-09-724-676A-70781
; Sequence 70781, Application US/09724676A
```

```

; GENERAL INFORMATION:
; APPLICANT: CompuGen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 CompuGen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 70781
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-70781

```

```

Query Match      23.7%; Score 59; DB 5; Length 555;
Best Local Similarity 47.6%; Pred. No. 2.5e+02;
Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 22 GCGPTPSNRAVKRKRPRSP 42
    |||||:|:|:|:|:|
Db 531 GCGPVPAGGARPHAPPPAAP 551

```

Search completed: April 2, 2003, 16:04:43  
Job time : 5.82192 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 2, 2003, 15:50:19 ; Search time 1.60731 Seconds  
(without alignments)  
2631.680 Million cell updates/sec

Title: US-09-079-678-51

Perfect score: 249

Sequence: 1 REFARRLWGCDLSRWRLDA.....PTPSNRAVVKRKRPRSPAL 44

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_73:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	67.5	27.1	486	2 G70579	probable murD prote
2	58	23.3	171	2 T49538	hypothetical prote
3	58	23.3	439	2 I60916	HNF-3/forkhead hom
4	58	23.3	456	2 T35474	50KD proline rich
5	56.5	22.7	435	2 I54182	tumor necrosis fac
6	56	22.5	472	2 A33312	sarcoplasmic retic
7	56	22.5	660	2 H71376	probable tpr prote
8	56	22.5	908	2 A33280	sarcolumenin precu
9	55.5	22.3	441	2 F98306	agaE protein [impo
10	55.5	22.3	441	2 AE2976	agaE protein [impo
11	55.5	22.3	490	2 B87023	UDP-N-acetylmuramo
12	55	22.1	286	2 A55530	megakaryocyte grow
13	55	22.1	519	2 D82536	conserved hypothet
14	55	22.1	1291	2 T00019	period protein hom
15	54	21.7	86	2 T23303	hypothetical prote
16	54	21.7	90	2 T23304	hypothetical prote
17	54	21.7	101	2 D83375	hypothetical prote
18	53.5	21.5	340	2 S04141	chitinase (EC 3.2.
19	53.5	21.5	347	2 S00549	developmental cont
20	53	21.3	168	2 T07640	pEARL1.1 protein -
21	53	21.3	356	2 A36070	L-lactate dehydrog
22	53	21.3	366	2 G82970	probable ferredoxi
23	53	21.3	2110	2 H96803	unknown protein t5
24	53	21.3	2251	2 T24490	hypothetical prote
25	52.5	21.1	491	2 S45068	53K glycoprotein -
26	52.5	21.1	593	2 S45025	glycoprotein G - s
27	52.5	21.1	999	2 T36021	probable zinc-bind
28	52.5	21.1	2512	2 E70751	probable nrp prote
29	52	20.9	420	2 T46910	hypothetical prote

ALIGNMENTS

RESULT 1

G70579

probable murD protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000

C:Accession: G70579

R: Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, N.; Hamlin, N.; Holroyd,

Rajandream, M. A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A: Authors: Squares, R.; Sulston, J. E.; Taylor, K.; Whitehead, S.; Barrell, B. G.

A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A: Reference number: A70500; MUID: 98295987; PMID: 9634230

A: Accession: G70579

A: Status: preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-486 <COL>

A: Cross-references: GB: Z95388; GB: AL123456; NID: g3261759; PIDN: CAB08672.1; PID: g21043

A: Experimental source: strain H37RV

C: Genetics:

C: Gene: murD

C: Superfamily: UDP-N-acetylmuramate-alanine ligase

Query Match	27.1%	Score	67.5;	DB	2;	Length	486;
Best Local Similarity	59.1%	Pred. No.	1.1;				
Matches	13;	Conservative	3;	Mismatches	5;	Indels	1;
Gaps	1;						

QY 8 LMGCDDLRLDAEGC-GTPPS 28

:|| :||||||| |||

Db 93 IWGDVELAWRLDAAGCYGPPRS 114

RESULT 2

T49538

hypothetical protein B21J21.170 [imported] - Neurospora crassa

C: Species: Neurospora crassa

C: Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 18-Aug-2000

C: Accession: T49538

R: Schultze, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu

submitted to the Protein Sequence Database, May 2000

A: Reference number: Z25022

A: Accession: T49538

A: Status: preliminary

A: Molecule type: DNA

A: Residues: 1-171 <SCH>

A: Cross-references: EMBL: AL355929; GSPDB: GN00116; NCSP: B21J21.170

A: Experimental source: BAC clone B21J21; strain OR74A

C: Genetics:

A: Gene: NCSP: B21J21.170

A: Map position: 6

C: Superfamily: Neurospora crassa hypothetical protein B21J21.170

C>Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 17-Mar-2000

C/Accession: I54182

C/RBaens, M.; Chafraet, M.; Cassiman, J.J.; Van den Berghe, H.; Marynen, P. Genomics 16, 214-218, 1993

A>Title: Construction and evaluation of a hncDNA library of human 12p transcribed seq

A/Reference number: I54182; MUID:93252381; PMID:8486360

A/Accession: I54182

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-435 <RES>

A/Cross-references: GB:L04270; NID:g339761; PIDN:AAA36757.1; PID:g339762

C/Genetics:

A/Gene: GDB:LTBR

A/Cross-references: GDB:I1230195; OMIM:600979

A/Map position: 12p13.3-12p13.1

C/Superfamily: tumor necrosis factor receptor type 1; NGF receptor repeat homology

Query Match 22.7%; Score 56.5; DB 2; Length 435;  
Best Local Similarity 44.0%; Pred. No. 23;  
Matches 11; Conservative 5; Mismatches 8; Indels 1; Gaps 1;

QY 12 DLSWRL-DAEGCGTPPSNRAVKHR 35  
          | : | : | : | | | | | : : :  
DB 406 DGKAHHLAETEHCGATPSNRCPNQ 430

RESULT 6

A33312

sarcoplasmic reticulum 53K glycoprotein precursor - rabbit

C/Species: Oryctolagus cuniculus (domestic rabbit)

C/Date: 08-Dec-1989 #sequence\_revision 08-Dec-1989 #text\_change 10-Dec-1999

C/Accession: A33312

R/Leberer, E.; Charuk, J.H.M.; Clarke, D.M.; Green, N.M.; Zubrzycka-Gaarn, E.; MacIen J. Biol. Chem. 264, 3484-3493, 1989

A>Title: Molecular cloning and expression of cDNA encoding the 53,000-Dalton glycoprotein

A/Reference number: A33312; MUID:89123480; PMID:2521635

A/Accession: A33312

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-472 <LEB>

A/Cross-references: GB:J04480; NID:g623544; PIDN:AAA60730.1; PID:g623545

C/Superfamily: sarcoplasmic reticulum 53K glycoprotein

C/Keywords: alternative splicing; glycoprotein

Query Match 22.5%; Score 56; DB 2; Length 472;  
Best Local Similarity 62.5%; Pred. No. 29;  
Matches 10; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 19 DAEGCGTPPSNRAVKH 34  
          | | | | | | | | | |  
DB 457 DKTGCGETPKNRKKH 472

RESULT 7

H71376

probable tpr protein B (tprB) - syphilis spirochete

C/Species: Treponema pallidum subsp. pallidum (syphilis spirochete)

C/Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 21-Jan-2000

C/Accession: H71376

R/Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gerson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; Mithy, L.; Weidman, J.; Smith, H.O.; Venter, J.C. Science 281, 375-388, 1998

A>Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.

A/Reference number: A71250; MUID:98332770; PMID:9665876

A/Accession: H71376

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-660 <COL>

A/Cross-references: GB:AE001186; GB:AE000520; NID:g3322263; PIDN:AAAC65006.1; PID:g3322263

A/Experimental source: strain Nichols

C/Genetics:

A/Gene: TP0011



## C:Genetics:

A:Gene: MGDF

A:Map position: 3q26.3

C:Keywords: alternative splicing; cytokine

Query Match 22.1%; Score 55; DB 2; Length 286;

Best Local Similarity 34.3%; Pred. No. 24;

Matches 12; Conservative 4; Mismatches 19; Indels 0; Gaps 0;

QY 9 WGCDDLWRLDAEGCGPTPSNRAVKHRKPRSP 43

Db 217 WNSWTLSWLTQDPRSPGFLNRIRLRATQPPA 251

## RESULT 13

D82536

conserved hypothetical protein XF2619 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 01-Mar-2002

C:Accession: D82536

R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717; PMID:10910347

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: D82536

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-519 &lt;SIM&gt;

A:Cross-references: GB:AE004068; GB:AE003849; NID:g9107832; PIDN:AAF85416.1; GSPDB:GN001

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H

as-Neto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm

Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir

M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF2619

C:Superfamily: hypothetical protein b1706

Query Match 22.1%; Score 55; DB 2; Length 519;

Best Local Similarity 39.3%; Pred. No. 42;

Matches 11; Conservative 4; Mismatches 13; Indels 0; Gaps 0;

QY 8 LWCDDLWRLDAEGCGPTPSNRAVKHR 35

Db 110 LLGADGVVYELQKGAGTPYSRGADGR 137

## RESULT 14

T00019

period protein homolog - mouse

C:Species: Mus musculus (house mouse)

C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 21-Jul-2000

C:Accession: T00019

R:Tei, H.; Okamura, H.; Shigeyoshi, Y.; Fukuhara, C.; Ozawa, R.; Hirose, M.; Sakaki, Y.

Nature 389, 512-516, 1997

A:Title: Circadian oscillation of a mammalian homologue of the Drosophila period gene.

A:Reference number: Z14056; MUID:97472418; PMID:9333243

A:Accession: T00019

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1291 &lt;TEI&gt;

A:Cross-references: EMBL:AB002108; NID:g2506046; PIDN:BAA22634.1; PID:g2506047

A:Experimental source: adult brain

C:Genetics:

A:Gene: mPer

A:Map position: 11B

Query Match 22.1%; Score 55; DB 2; Length 1291;

Best Local Similarity 34.0%; Pred. No. 1e+02;

Matches 17; Conservative 3; Mismatches 16; Indels 14; Gaps 2;

QY 7 RLWGCDDLWRLDAEGC--GPTP-----SNRAVKHRKPRSP 42

Db 803 RLRLGDTSSVAFSPGCHHGIPPPGRRHCRSKAKSRHHHHOTPRP 852

## RESULT 15

T25303

hypothetical protein T26E4.9 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T25303

R:McMurray, A.

submitted to the EMBL Data Library, October 1996

A:Reference number: Z20013

A:Accession: T25303

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-86 &lt;WIL&gt;

A:Cross-references: EMBL:Z81132; PIDN:CAB03428.1; GSPDB:GN00023; CESP:T26E4.9

A:Experimental source: clone T26E4

C:Genetics:

A:Gene: CESP:T26E4.9

A:Map position: 5

A:Introns: 26/3

Query Match 21.7%; Score 54; DB 2; Length 86;

Best Local Similarity 38.2%; Pred. No. 9.7;

Matches 13; Conservative 1; Mismatches 12; Indels 8; Gaps 2;

QY 6 RRLWGCDDLWRLDAEGCGPTPSNRAVKHRKPRP 39

Db 49 KRQWGGCDCWGC---CGTPDPAAAT---PAP 74

Search completed: April 2, 2003, 15:54:29

Job time : 3.60731 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 2, 2003, 15:48:04 ; Search time 0.90411 seconds  
(without alignments)  
2018.515 Million cell updates/sec

Title: US-09-079-678-51  
Perfect score: 249  
Sequence: 1 REFAERRLGCGDLSRLDA.....PTPSNRVAKHRKPRSPAL 44

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	67.5	27.1	486	1 MURD_MYCTU	O06222 mycobacteri
2	58	23.3	439	1 HPFH_RAT	O63444 rattus norv
3	56.5	22.7	435	1 TNK3_HUMAN	P36941 homo sapien
4	56	22.5	908	1 SRCA_RABIT	P13666 oryctolagus
5	55.5	22.3	490	1 MURD_MYCLE	P57995 mycobacteri
6	55	22.1	870	1 DYN2_MOUSE	P39054 mus musculu
7	55	22.1	1291	1 PER1_MOUSE	O35973 mus musculu
8	53.5	21.5	347	1 ZFP2_MOUSE	P08043 mus musculu
9	53	21.3	356	1 LDHA_HORVU	P22988 hordeum vul
10	53	21.3	562	1 GUN1_ACICE	P54583 acidothermu
11	52	20.9	870	1 DYN2_HUMAN	P50570 homo sapien
12	52	20.9	1559	1 STCJ_EMENI	O00681 emericella
13	51.5	20.7	236	1 GTX2_WAIZE	P50472 zea mays (m
14	51.5	20.7	323	1 CGS6_HUMAN	Q9Y3A4 homo sapien
15	51.5	20.7	724	1 MALQ_MYCTU	O53932 mycobacteri
16	51.5	20.7	835	1 AXN1_BRARE	P57094 brachydanio
17	51.5	20.7	968	1 BCAL_RAT	O63767 rattus norv
18	51.5	20.7	1995	1 YCX7_CHLRE	P36495 chlamydomon
19	51	20.5	148	1 YLK2_EBVAB	P07285 epstein-bar
20	51	20.5	199	1 EQST_ACTEQ	P81439 actinia equ
21	51	20.5	566	1 TS13_MOUSE	Q01755 mus musculu
22	51	20.5	658	1 SQHC_BRAJA	P54924 bradyrhizob
23	51	20.5	904	1 VGLB_HSV11	P10211 herpes simp
24	51	20.5	2769	1 THYG_BOVIN	P01267 bos taurus
25	51	20.5	3312	1 CLR3_HUMAN	Q9N9V7 homo sapien
26	50	20.1	118	1 LSHB_PHYCA	P25330 physeter ca
27	50	20.1	451	1 CSSS_BACSU	O32193 bacillus su
28	50	20.1	598	1 TFDB_ALCEU	P27138 alcaligenes
29	50	20.1	647	1 SOHC_RHISN	P55348 rhizobium s
30	50	20.1	737	1 PEN_DROME	O61345 drosophila
31	50	20.1	878	1 VAV2_HUMAN	P52735 homo sapien
32	50	20.1	1180	1 VYK2_MOUSE	Q9R117 mus musculu
33	49.5	19.9	359	1 RFBG_SALTY	P26397 salmonella

34	49	19.7	118	1 LSHB_BALAC	P33088 balaenopter
35	49	19.7	206	1 RNFG_ECO57	P58345 escherichia
36	49	19.7	206	1 RNFG_ECOLI	P77285 escherichia
37	49	19.7	250	1 UNG_HCAVA	P16769 human cytom
38	49	19.7	405	1 VPAP_HSVB	P28958 equine herp
39	49	19.7	477	1 XYNA_STRLI	P26514 streptomyce
40	49	19.7	558	1 C4BP_RAT	O63514 rattus norv
41	49	19.7	575	1 YFBK_ECOLI	P76481 escherichia
42	49	19.7	777	1 UNG8_CAEEL	Q21974 caenorhabdi
43	49	19.7	870	1 DYN2_RAT	P39052 rattus norv
44	49	19.7	2076	1 PAS1_YARLI	P34229 y fatty aci
45	49	19.7	2768	1 THYG_HUMAN	P01266 homo sapien

ALIGNMENTS

RESULT 1					
MURD_MYCTU					
ID	MURD_MYCTU	STANDARD;	PRT;	486 AA.	
AC	O06222:				
DT	15-JUL-1998 (Rel. 36, Created)				
DT	15-JUL-1998 (Rel. 36, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	UDP-N-acetylmuramoyl-L-alanine--D-glutamate ligase (EC 6.3.2.9) (UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase) (D-glutamic acid adding enzyme).				
DE	acetyluramoyl-L-alanyl-D-glutamate synthetase)				
GN	MURD OR RV2155C OR MT2214 OR MTCY270.13.				
OS	Mycobacterium tuberculosis.				
OC	Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;				
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.				
OX	NCBI_TaxID=1773;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=H37RV;				
RX	MEDLINE=98295987; PubMed=9634230;				
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;				
RA	"Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence."				
RT	Nature 393:537-544(1998).				
RL	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CDC 1551 / Oshkosh;				
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O., Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E., Kolonay J.F., Nelson W.C., Umayam L.A., Emdolaeva M.D., Salzberg S.L., Delcher A., Utterback T., Weldman J., Khouri H., Gill J., Mikula A., Bishai W.;				
RA	"Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains."				
RT	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.				
RL	-!- FUNCTION: CELL WALL FORMATION. CATALYZES THE ADDITION OF D-GLUTAMATE TO THE NUCLEOTIDE PRECURSOR UDP-N-ACETYLMURAMOYL-L-ALANINE (UMA) (BY SIMILARITY).				
CC	-!- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramoyl-L-alanine + D-glutamate -> ADP + phosphate + UDP-N-acetylmuramoyl-L-alanyl-D-glutamate.				
CC	-!- PATHWAY: peptidoglycan biosynthesis.				
CC	-!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).				
CC	-!- SIMILARITY: BELONGS TO THE MURDEF FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial				

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC  
CC  
CC  
DR EMBL; Z95388; CAB08672.1; -;  
DR EMBL; AE007068; AAK46498.1; ALT\_INIT.  
DR HSSP; P14900; 1E0D.  
DR TIGR; MT2214; -;  
DR Tuberculin; RV2155c; -;  
DR InterPro; IPR000713; Mur\_ligase.  
DR InterPro; IPR004101; Mur\_ligase\_C.  
DR Pfam; PF01225; Mur\_ligase; 1.  
DR Pfam; PF02875; Mur\_ligase\_C; 1.  
DR TIGR; TIGR01087; murD; 1.  
DR Peptidoglycan synthetase; Cell wall; Cell division; Ligase;  
KW ATP-binding; Complete proteome.  
FT NP\_BIND 120 126 ATP (POTENTIAL).  
FT DOMAIN 84 89 POLY-ALA.  
FT DOMAIN 125 128 POLY-THR.  
FT DOMAIN 284 291 POLY-ALA.  
FT CONFLICT 80 80 T -> I (IN REF. 2).  
FT CONFLICT 247 247 R -> G (IN REF. 2).  
SQ SEQUENCE 486 AA; 49345 MW; 5C4BBCD2AF5FAC95 CRC64;  
  
Query Match 27.1%; Score 67.5; DB 1; Length 486;  
Best Local Similarity 59.1%; Pred. No. 0.23;  
Matches 13; Conservative 3; Mismatches 5; Indels 1; Gaps 1;  
  
QY 8 LWGCDLWRLDAEGC-GPTPS 28  
DB 93 IINGDVELAWRLDAAGCYGPPRS 114  
  
RESULT 2  
HFH1\_RAT  
ID HFH1\_RAT STANDARD; PRT; 439 AA.  
AC Q63244;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Hepatocyte nuclear factor 3 forkhead homolog 1 (HFH-1).  
GN HFH1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Sprague-Dawley; TISSUE=Lung;  
RX MEDLINE=93248207; PubMed=7683413;  
RA Clevidence D.E., Overdier D.G., Tao W., Qian X., Pani L., Lai E., Costa R.H.;  
RT "Identification of nine tissue-specific transcription factors of the hepatocyte nuclear factor 3/forkhead DNA-binding domain family.";  
RL Proc. Natl. Acad. Sci. U.S.A. 90:3948-3952(1993).  
CC -!- SUBCELLULAR LOCATION: Nucleus.  
CC -!- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR PROSITE; PS00657; FORK\_HEAD\_1; 1.  
DR PROSITE; PS00658; FORK\_HEAD\_2; 1.  
DR PROSITE; PS00039; FORK\_HEAD\_3; 1.  
KW DNA-binding; Nuclear protein; Transcription regulation.  
FT DNA\_BIND 99 190  
SQ SEQUENCE 439 AA; 45481 MW; A37028448644C350 CRC64;  
  
Query Match 23.3%; Score 58; DB 1; Length 439;  
Best Local Similarity 36.8%; Pred. No. 3.6;  
Matches 14; Conservative 2; Mismatches 14; Indels 8; Gaps 1;  
  
QY 7 RLWGCDDLWRLDAEGCPTPSNRAVVKHKKPRRSPAL 44  
DB 270 RLWGCSPYGGALLPARRCAPIP-----RSFPRRPAV 299  
  
RESULT 3  
TNR3\_HUMAN  
ID TNR3\_HUMAN STANDARD; PRT; 435 AA.  
AC P36941;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Tumor necrosis factor receptor superfamily member 3 precursor (lymphotoxin-beta receptor) (Tumor necrosis factor receptor 2 related protein) (Tumor necrosis factor C receptor).  
DE LTBR OR TNFRSF3 OR TNFCR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=93252381; PubMed=8486360;  
RA Baens M., Chaffanet M., Cassiman J.J., den Berghe H., Marynen P.;  
RT "Construction and evaluation of a hncDNA library of human l2p transcribed sequences derived from a somatic cell hybrid.";  
RL Genomics 16:214-218(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RA Strausberg R.;  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP FUNCTION  
RX MEDLINE=94225209; PubMed=8171323;  
RA Crowe P.D., van Arsdale T.L., Walter B.N., Ware C.F., Hession C., Ehrenfels B., Browning J.L., Din W.S., Goodwin R.G., Smith C.A.;  
RT "A lymphotoxin-beta-specific receptor.";  
RL Science 264:707-710(1994).  
RN [4]  
RP CHARACTERIZATION.  
RX MEDLINE=99223511; PubMed=10207006;  
RA Wu M.-Y., Wang P.-Y., Han S.-H., Hsieh S.-L.;  
RT "The cytoplasmic domain of the lymphotoxin-beta receptor mediates cell death in HeLa cells.";  
RL J. Biol. Chem. 274:11868-11873(1999).  
RN [5]  
RP FUNCTION.  
RX MEDLINE=20261554; PubMed=10799510;  
RA Rooney I.A., Butrovich K.D., Glass A.A., Borboroglu S., Benedict C.A., Whitbeck J.C., Cohen G.H., Eisenberg R.J., Ware C.F.;  
RT "The lymphotoxin-beta receptor is necessary and sufficient for LIGHT-mediated apoptosis of tumor cells.";  
RL J. Biol. Chem. 275:14307-14315(2000).  
CC -!- FUNCTION: Receptor for the heterotrimeric lymphotoxin containing LTA and LTb, and for TNFSF14/LIGHT. Promotes apoptosis via TRAF3 and TRAF5. May play a role in the development of lymphoid organs.  
CC -!- SUBUNIT: Self-associates.  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- SIMILARITY: CONTAINS 4 TNFR-CYS REPEATS.



RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
RA Barrell B.G.;  
RT "Massive gene decay in the leprosy bacillus."  
RL Nature 409:1007-1011(2001).  
CC -!- FUNCTION: CELL WALL FORMATION. CATALYZES THE ADDITION OF D-  
CC GLUTAMATE TO THE NUCLEOTIDE PRECURSOR UDP-N-ACETYLMURAMOYL-L-  
CC ALANINE (UMA) (BY SIMILARITY).  
CC CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramoyl-L-alanine + D-  
CC glutamate -> ADP + phosphate + UDP-N-acetylmuramoyl-L-alanyl-D-  
CC glutamate.  
CC -!- PATHWAY: Peptidoglycan biosynthesis.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -!- SIMILARITY: BELONGS TO THE MURCEDEF FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AL583920; CAC31293.1; -;  
DR EMBL; AL022602; CAAL8670.1; -;  
DR HSSP; P14900; 1UAG.  
DR Leproma; ML0912; -;  
DR InterPro; IPR000713; Mur\_ligase.  
DR InterPro; IPR004101; Mur\_ligase\_C.  
DR Pfam; PF01225; Mur\_ligase; 1.  
DR Pfam; PF02875; Mur\_ligase\_C; 1.  
DR TIGRFAMs; TIGR01087; murD; 1.  
KW Peptidoglycan synthesis; Cell wall; Cell division; Ligase;  
KW ATP-binding; Complete proteome.  
FT NP\_BIND 124 130 ATP (POTENTIAL).  
FT FT  
SQ SEQUENCE 490 AA; 50583 MW; 1385495843830536 CRC64;  
Query Match 22.3%; Score 55.5; DB 1; Length 490;  
Best Local Similarity 57.9%; Pred. No. 8.5;  
Matches 11; Conservative 3; Mismatches 4; Indels 1; Gaps 1;  
OY 8 LWGCDLWSRLDAEG-CGP 25  
Db :||:|||||:|  
97 IWGDVELAWRLDAAGYGP 115  
RESULT 6  
DYN2\_MOUSE STANDARD; PRT; 870 AA.  
AC P39054; Q9DBEL;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE DYNamin 2 (EC 3.6.1.50) (DYNamin UDNM).  
GN DNM2 OR DYN2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 2).  
RC STRAIN=NIH Swiss;  
RX MEDLINE=97288532; PubMed=9143510;  
RA Klocke R., Augustin A., Ronsiek M., Stief A., van der Putten H.,  
RA Jockusch H.;  
RT "DYNamin genes Dnm1 and Dnm2 are located on proximal mouse  
RT chromosomes 2 and 9, respectively."  
RL Genomics 41:290-292(1997).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC STRAIN=C57BL/6J; TISSUE=Liver;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,

RA Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staebli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection."  
RL Nature 409:685-690(2001).  
CC -!- FUNCTION: Microtubule-associated force-producing protein involved  
CC in producing microtubule bundles and able to bind and hydrolyze  
CC GTP. Most probably involved in vesicular trafficking processes, in  
CC particular endocytosis.  
CC -!- CATALYTIC ACTIVITY: GTP + H(2)O -> GDP + phosphate.  
CC -!- SUBUNIT: Interacts with SHANK1 and SHANK2 (By similarity). Also  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic. Microtubule-associated. Also  
CC found in the postsynaptic density of neuronal cells (By  
CC similarity).  
CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are  
CC produced by alternative splicing.  
CC -!- SIMILARITY: BELONGS TO THE DYNAMIN FAMILY.  
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; L31398; AAA40523.1; -;  
DR EMBL; AK005012; BAB23745.1; -;  
DR HSSP; Q05193; 1DYN.  
DR MGD; MGI:109547; Dnm2.  
DR InterPro; IPR001401; DYNamin.  
DR InterPro; IPR000375; DYNamin\_central.  
DR InterPro; IPR003130; GED.  
DR InterPro; IPR001849; PH.  
DR Pfam; PF00169; PH; 1.  
DR Pfam; PF00350; dymamin; 1.  
DR Pfam; PF01031; dymamin\_2; 1.  
DR Pfam; PF02212; GED; 1.  
DR PRINTS; PR00195; DYNAMIN.  
DR SMART; SM00053; DYNC; 1.  
DR SMART; SM00302; GED; 1.  
DR SMART; SM00233; PH; 1.  
DR PROSITE; PS00410; DYNAMIN; 1.  
DR PROSITE; PS50003; PH\_DOMAIN; 1.  
DR Hydrolase; Motor protein; GTP-binding; Microtubules; Multigene family;  
KW Endocytosis; Alternative splicing.  
FT NP\_BIND 38 45 GTP (BY SIMILARITY).  
FT NP\_BIND 136 140 GTP (BY SIMILARITY).  
FT NP\_BIND 205 208 GTP (BY SIMILARITY).  
FT DOMAIN 519 625 PH.  
FT VARSPIC 516 519 MISSING (IN ISOFORM 2).  
FT CONFLICT 297 298 RS -> HG (IN REF. 1).  
FT CONFLICT 848 870 SRRAPADSRPIIRPAESLLD -> RRPPLAPARPPF  
FT (IN REF. 2).  
SQ SEQUENCE 870 AA; 98145 MW; E80864AF94B8F778 CRC64;  
Query Match 22.1%; Score 55; DB 1; Length 870;  
Best Local Similarity 37.5%; Pred. No. 18;  
Matches 12; Conservative 3; Mismatches 17; Indels 0; Gaps 0;





```
CC DIFFERENTIATION AND/OR PHENOTYPIC MAINTENANCE OF NEURONS.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y00850; CAA68768.1; -.
DR EMBL; M15709; AAA37640.1; -.
DR PIR; S00549; S00549.
DR PIR; B26595; B26595.
DR HSSP; P08046; 1A1H.
DR MGD; MGI:99167; Zfp2.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; Zf_C2H2; 9.
DR PRINTS; PR00048; ZINCFINGER.
DR ProDom; PD000003; Znf_C2H2; 9.
DR SMART; SM00355; Znf_C2H2; 9.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 9.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 9.
KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW Repeat; Nuclear protein.
FT DOMAIN 100 346 ZINC FINGERS.
FT ZN_FING 100 122 C2H2-TYPE.
FT ZN_FING 128 150 C2H2-TYPE.
FT ZN_FING 156 178 C2H2-TYPE.
FT ZN_FING 184 206 C2H2-TYPE.
FT ZN_FING 212 234 C2H2-TYPE.
FT ZN_FING 240 262 C2H2-TYPE.
FT ZN_FING 268 290 C2H2-TYPE.
FT ZN_FING 296 318 C2H2-TYPE.
FT ZN_FING 324 346 C2H2-TYPE.
SQ SEQUENCE 347 AA; 39855 MW; 4F889F6CDF548C CRC64;

Query Match 21.5%; Score 53.5; DB 1; Length 347;
Best Local Similarity 28.9%; Pred No. 11;
Matches 11; Conservative 10; Mismatches 6; Indels 11; Gaps 2;

QY 1 REFAERLWGCDDLWRLDAECGPT--PSNRAVKHRK 36
| | : : : : | | | | | | | | : : | | :
| | : : : : | | | | | | | | : : | | :
Db 92 RMFQKKVYGCDE-----CGTKFRSSSLKKHQ 120

RESULT 9
LDHA_HORVU
ID LDHA_HORVU STANDARD; PRT; 356 AA.
AC P22988;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE L-lactate dehydrogenase A (EC 1.1.1.27) (LDH-A).
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=cv. Himalaya; TISSUE=Root;
RX MEDLINE=90384994; PubMed=1598294;
RT Hondred D.; Hanson A.D.;
RA "Hypoxically inducible barley lactate dehydrogenase: cDNA cloning and
RT molecular analysis."
RL Proc. Natl. Acad. Sci. U.S.A. 87:7300-7304(1990).
CC -!- CATALYTIC ACTIVITY: (S)-lactate + NAD(+) = pyruvate + NADH.
CC -!- PATHWAY: Anaerobic glycolysis; final step.
CC -!- SUBUNIT: Tetramer that arise from random association of LDH-A
```

```
CC AND LDH-B.
CC -!- INDUCTION: BY HYPOXIA.
CC -!- SIMILARITY: BELONGS TO THE LDH FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M55685; AAA62696.1; -.
DR PIR; A36070; A36070.
DR HSSP; P00342; 2LDX.
DR InterPro; IPR001557; L_LDH.
DR InterPro; IPR000205; NAD_binding.
DR InterPro; IPR001236; ldh.
DR Pfam; PF00056; ldh; 1.
DR Pfam; PF02866; ldh_C; 1.
DR PRINTS; PS00086; LLDHGRNASE.
DR PROSITE; PS00064; L_LDH; 1.
KW Oxidoreductase; NAD; Glycolysis; Multigene family.
FT ACT_SITE 215 215 ACCEPTS A PROTON DURING CATALYSIS
(BY SIMILARITY).
SQ SEQUENCE 356 AA; 37749 MW; 7B82D6F1D0BF5AD8 CRC64;

Query Match 21.3%; Score 53; DB 1; Length 356;
Best Local Similarity 52.9%; Pred No. 13;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 20 AEGCGPTPSNRAVKHRK 36
| : | | | : | | | |
Db 26 ADGCPATPTSSAVPHRR 42

RESULT 10
GUNL_ACICE
ID GUNL_ACICE STANDARD; PRT; 562 AA.
AC P54583;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Endoglucanase E1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase E1)
DE (Cellulase E1) (Endocellulase E1).
OS Acidothermus cellulolyticus.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Frankineae; Acidothermaceae; Acidothermus.
OX NCBI_TaxID=28049;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 43068 / 11B;
RA Laymon R.A.; Himmel M.E.; Thomas S.R.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 42-398.
RX MEDLINE=96346058; PubMed=8718854;
RA Sakon J.; Adney W.S.; Himmel M.E.; Thomas S.R.; Kaplus P.A.;
RT "Crystal structure of thermostable family 5 endocellulase E1 from
RT Acidothermus cellulolyticus in complex with cellotetraose."
RL Biochemistry 35:10648-10660(1996).
CC -!- FUNCTION: THERMOSTABLE ENZYME WITH AN OPTIMAL TEMPERATURE OF 81
CC DEGREES CELSIUS. HAS A VERY HIGH SPECIFIC ACTIVITY ON
CC CARBOXYMETHYLCELLULOSE.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: U33212; AAA75477.1; -  
 CC PDB: 1BCE; 14-OCT-96.  
 CC InterPro: IPR001919; Bac\_celose-bind.  
 CC InterPro: IPR001547; GH\_5.  
 CC Pfam: PF00150; cellulase; 1.  
 CC Pfam: PF00553; CBM\_2; 1.  
 CC PROSITE: PS00659; GLYCOSYL-HYDROL\_F5; 1.  
 KW Cellulose degradation; Hydrolase; Glycosidase; Signal; 3D-structure.  
 FT SIGNAL 1 41  
 FT CHAIN 42 562  
 FT DOMAIN 42 400  
 FT DOMAIN 401 461  
 FT DOMAIN 462 562  
 FT ACT\_SITE 203 203  
 FT ACT\_SITE 323 323  
 FT DISULFID 75 161  
 FT DISULFID 209 212  
 SQ SEQUENCE 562 AA; 60747 MW; 84E6256406A35041 CRC64;

Query Match 21.3%; Score 53; DB 1; Length 562;  
 Best Local Similarity 55.6%; Pred. No. 21;  
 Matches 10; Conservative 1; Mismatches 5; Indels 2; Gaps 1;  
 QY 9 WCCDD-LSWRLDARECG 24  
 ||||| : ||||| |  
 Db 210 WCCGDPISDWRLAERAG 227

RESULT 11  
 DYN2\_HUMAN  
 ID DYN2\_HUMAN STANDARD; PRT; 870 AA.  
 AC P50570;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Dynamin 2 (EC 3.6.1.50).  
 GN DNM2 OR DYN2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RX MEDLINE=96011652; PubMed=7590285;  
 RA Diatloff-Wito C., Gordon A.J.E., Duchaud E., Merlin G.;  
 RT "Isolation of an ubiquitously expressed cDNA encoding human dynamin  
 RL II, a member of the large GTP-binding protein family.";  
 RL Gene 163:301-306(1995).  
 RN [2]  
 RP INTERACTION WITH SHANK PROTEINS.  
 RX PubMed=11583995;  
 RA Okamoto P.M., Gamby C., Wells D., Fallon J., Vallee R.B.;  
 RT "Dynamin isoform-specific interaction with the shank/prosAP  
 RT scaffolding proteins of the postsynaptic density and actin  
 RT cytoskeleton.";  
 RL J. Biol. Chem. 276:48458-48465(2001).  
 CC -!- FUNCTION: MICROTUBULE-ASSOCIATED FORCE-PRODUCING PROTEIN INVOLVED  
 CC IN PRODUCING MICROTUBULE BUNDLES AND ABLE TO BIND AND HYDROLYZE  
 CC GTP. MOST PROBABLY INVOLVED IN VESICULAR TRAFFICKING PROCESSES, IN  
 CC PARTICULAR ENDOCYTOSIS.  
 CC -!- CATALYTIC ACTIVITY: GTP + H(2)O = GDP + phosphate.  
 CC -!- SUBUNIT: Interacts with SHANK1 and SHANK2.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic. Microtubule-associated. Also  
 CC found in the postsynaptic density of neuronal cells.  
 CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are  
 CC produced by alternative splicing.  
 CC -!- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED.  
 CC -!- SIMILARITY: BELONGS TO THE DYNAMIN FAMILY.  
 CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: L36983; AAA88025.1; -  
 CC HSP: Q05193; 2DYN.  
 CC Genew; HGNC:2974; DNM2.  
 CC MIM; 602378; -  
 CC InterPro: IPR001401; Dynamin.  
 CC InterPro: IPR000375; Dynamin\_central.  
 CC InterPro: IPR003130; GED.  
 CC InterPro: IPR001849; PH.  
 CC Pfam: PF00169; PH; 1.  
 CC Pfam: PF00350; dynamin; 1.  
 CC Pfam: PF01031; dynamin\_2; 1.  
 CC Pfam: PF02212; GED; 1.  
 CC PRINTS: PR00195; DYNAMIN.  
 CC SMART; SM00053; DYNC; 1.  
 CC SMART; SM00302; GED; 1.  
 CC SMART; SM00233; PH; 1.  
 CC PROSITE: PS00410; DYNAMIN; 1.  
 CC PROSITE: PS00003; PH\_DOMAIN; 1.  
 KW Hydrolase; Motor protein; GTP-binding; Microtubules; Multigene family;  
 KW Endocytosis; Alternative splicing.  
 FT NP\_BIND 38 45 GTP (BY SIMILARITY).  
 FT NP\_BIND 136 140 GTP (BY SIMILARITY).  
 FT NP\_BIND 205 208 GTP (BY SIMILARITY).  
 FT DOMAIN 519 625 PH.  
 FT VARSPPLIC 516 519 MISSING (IN ISOFORM 2).  
 SQ SEQUENCE 870 AA; 98018 MW; 149189A598BE7039 CRC64;

Query Match 20.9%; Score 52; DB 1; Length 870;  
 Best Local Similarity 37.5%; Pred. No. 44;  
 Matches 12; Conservative 2; Mismatches 18; Indels 0; Gaps 0;  
 QY 13 DLSWRLDARECGPTPSNRAVKHKRPPSPAL 44  
 | : | | | | | | | | | | | | | |  
 Db 753 DDTWLQSSASHPTQRRPVSVSIHPPRPAPV 784

RESULT 12  
 STCJ\_EMENI  
 ID STCJ\_EMENI STANDARD; PRT; 1559 AA.  
 AC Q00681;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Putative sterigmatocystin biosynthesis fatty acid synthase alpha  
 DE subunit.  
 DE STCJ.  
 OS Emericella nidulans (Aspergillus nidulans).  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiales; Trichocomaceae; Emericella.  
 OX NCBI\_TaxID=5072;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FGSC 26;  
 RX MEDLINE=96202293; PubMed=8643646;  
 RA Brown D.W., Yu J.-H., Kelkar H.S., Fernandes M., Nesbitt T.C.,  
 RA Kellier N.P., Adams T.H., Leonard T.J.;  
 RT "Twenty-five coregulated transcripts define a sterigmatocystin gene  
 RT cluster in Aspergillus nidulans.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:1418-1422(1996).  
 CC -!- FUNCTION: INVOLVED IN THE SYNTHESIS OF STERIGMATOCYSTIN. MAY BE  
 CC INVOLVED IN THE SYNTHESIS OF A SIX-CARBON FATTY ACID THAT SERVES  
 CC TO INITIATE POLYKETIDE SYNTHESIS.  
 CC -!- PATHWAY: Sterigmatocystin biosynthesis.  
 CC -!- SIMILARITY: TO THE FATTY ACID SYNTHETASE, SUBUNIT ALPHA FROM

```

CC OTHER FUNCI.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U34740; AAC49198.1; -
DR InterPro: IPR002582; ACPS.
DR InterPro: IPR000794; Ketoacyl-synt.
DR InterPro: IPR004568; Pantethn.trn.
DR InterPro: IPR003880; Pantethn_attach.
DR Pfam: PF00109; ketoacyl-synt; 1.
DR Pfam: PF01648; ACPS; 1.
DR Pfam: PF02801; ketoacyl-synt_C; 1.
DR ProDom: PD004282; ACPS; 1.
DR TIGRFAMs: TIGR00556; pantethn.trn; 1.
DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; 1.
DR PROSITE: PS00606; B_KETOACYL_SYNTHASE; 1.
DR Fatty acid biosynthesis; Multifunctional enzyme; Oxidoreductase;
KW Transferrase; NADP; Phosphopantetheine.
FT DOMAIN 1 ? ACYL CARRIER (ACP).
FT DOMAIN 458 652 ? BETA-KETOACYL REDUCTASE.
FT DOMAIN ? ? BETA-KETOACYL SYNTHASE.
FT BINDING 106 106 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT ACT_SITE 1058 1058 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
SQ SEQUENCE 1559 AA; 170126 MW; F33D2245A2AE4430 CRC64;

Query Match 20.9%; Score 52; DB 1; Length 1559;
Best Local Similarity 38.1%; Pred. No. 82;
Matches 8; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 3 FAERLWGCDDLWRLDARGC 23
DB 790 FAELGFWGSARINWESRGC 810

RESULT 13
GTXX2_MAIZE
ID GTXX2_MAIZE STANDARD; PRT; 236 AA.
AC P50472;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable glutathione S-transferase BZ2 (EC 2.5.1.18) (Bronze-2
DE protein).
GN BZ2.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoidae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Seedling;
RX MEDLINE=93005645; PubMed=1967051;
RA Nash J., Luehrsen K.R., Walbot V.;
RT "Bronze-2 gene of maize: reconstruction of a wild-type allele and
RT analysis of transcription and splicing.";
RL Plant Cell 2:1039-1049(1990).
RN [2]
RP ERRATUM.
RA Nash J., Luehrsen K.R., Walbot V.;
RL Plant Cell 3:103-103(1991).
CC -!- CATALYTIC ACTIVITY: RX + glutathione = HX + S-glutathione.
CC -!- PATHWAY: Anthocyanin biosynthesis.
CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. HSP26 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U14599; AAA50245.1; -
DR MaizEDB: 64140; -
DR InterPro: IPR004045; GST_Nterm.
DR Pfam: PF02798; GST_N; 1.
KW Transferrase.
SQ SEQUENCE 236 AA; 25563 MW; BD7C087F60FA6E9A CRC64;

Query Match 20.7%; Score 51.5; DB 1; Length 236;
Best Local Similarity 45.5%; Pred. No. 13;
Matches 15; Conservative 0; Mismatches 15; Indels 3; Gaps 2;

QY 2 EFAERLWGC--DDLW--RLDARGCGGTPFSNRA 31
DB 95 ERAMHREFTAFIDDKFWPALDAVSLAPTFGARA 127

RESULT 14
CG96_HUMAN
ID CG96_HUMAN STANDARD; PRT; 323 AA.
AC Q9Y3A4; Q9Y3H5;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein CGI-96 (BK126B4.3).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20272150; PubMed=10810093;
RA Lai C.-H., Chou C.-Y., Chang L.-Y., Liu C.-S., Lin W.-C.;
RT "Identification of novel human genes evolutionarily conserved in
RT Caenorhabditis elegans by comparative proteomics.";
RL Genome Res. 10:703-713(2000).
RN [2]
RP SEQUENCE OF 46-323 FROM N.A.
RA Ellington A.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO C.ELEGANS ZC434.4 AND S.POMBE SPBC776.17.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF151854; AAD34091.1; -
DR EMBL: AL022316; CAB41546.1; -
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF00076; rtm; 1.
DR SMART; SM00360; RRM; 1.
KW Hypothetical protein.
FT CONFLICT 55 66
FT CONFLICT 237 237
FT CONFLICT 257 300
FT CONFLICT 237 237
FT CONFLICT 257 300
SQ SEQUENCE 323 AA; 35897 MW; C4CFABBC8B35FB82 CRC64;

Query Match 20.7%; Score 51.5; DB 1; Length 323;
Best Local Similarity 35.0%; Pred. No. 18;
Matches 14; Conservative 4; Mismatches 11; Indels 11; Gaps 2;

```



**THIS PAGE BLANK (uspto)**



Query Match	23.5%	Score 58.5;	DB 16;	Length 238;
Best Local Similarity	46.9%	Pred. No. 7.7;		
Matches 15; Conservative		3; Mismatches	9; Indels	5; Gaps
				1;

RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RX MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 RL the 8 Mb *Streptomyces coelicolor* A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).



```

RN  SEQUENCE FROM N.A.
RC  STRAIN-A3(2) / M145;
RA  Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA  Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA  Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA  Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA  Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA  Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA  Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA  Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA  Hopwood D.A.;
RT  "Complete genome sequence of the model actinomycete Streptomyces
RT  coelicolor A3(2).";
RL  Nature 417:141-147(2002).
DR  EMBL; AL034492; CAA22501.1; -.
DR  InterPro; IPR002965; P-rich_extensn.
DR  PRINTS; PR01217; PRICHEXTENS.
SQ  SEQUENCE 456 AA; 5032 MW; 1293F4A97256AC57 CRC64;

Query Match 23.3%; Score 58; DB 16; Length 456;
Best Local Similarity 52.6%; Pred. No. 18;
Matches 10; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 24 GPTPSNRVVKHKPRRSP 42
Db 120 GPAPVRAGHRSPRTGP 138

RESULT 6
Q8R0J3 PRELIMINARY; PRT; 794 AA.
AC Q8R0J3;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Similar to RIKEN CDNA 4122402022 gene.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC026767; AAH26767.1; -.
SQ SEQUENCE 794 AA; 89984 MW; 4B95263A2A253DD6 CRC64;

Query Match 22.9%; Score 57; DB 11; Length 794;
Best Local Similarity 42.4%; Pred. No. 43;
Matches 14; Conservative 3; Mismatches 14; Indels 2; Gaps 1;

QY 14 LSWRLDAEGCGTTP--SNRAVHKRPRRSPAL 44
Db 192 LSWLPTGSGVPPPLCSPPGSGSPSPAPRTPM 224

RESULT 7
Q9CX74 PRELIMINARY; PRT; 823 AA.
AC Q9CX74;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE 4122402022Rik protein.
GN 4122402022Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-C57BL/6J; TISSUE=EMBRYONIC HEAD;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yananaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kaniya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kotsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK019466; BAB31737.1; -.
DR MGD; MGI:1924876; 4122402022Rik.
SQ SEQUENCE 823 AA; 93305 MW; F96C8F7EBA227F10 CRC64;

Query Match 22.9%; Score 57; DB 11; Length 823;
Best Local Similarity 42.4%; Pred. No. 44;
Matches 14; Conservative 3; Mismatches 14; Indels 2; Gaps 1;

QY 14 LSWRLDAEGCGTTP--SNRAVHKRPRRSPAL 44
Db 192 LSWLPTGSGVPPPLCSPPGSGSPSPAPRTPM 224

RESULT 8
Q9ARU1 PRELIMINARY; PRT; 386 AA.
AC Q9ARU1;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)
DE P0445D12.6 protein.
GN P0445D12.6.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0445D12.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AP003046; BAB40035.1; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
SQ SEQUENCE 386 AA; 40066 MW; C9D71BF12E30EB45 CRC64;

Query Match 22.7%; Score 56.5; DB 10; Length 386;
Best Local Similarity 39.5%; Pred. No. 23;
Matches 17; Conservative 6; Mismatches 17; Indels 3; Gaps 3;

QY 1 REFAER-RLWGCDLRLSWRLDAEGCG-PTPSNRVHKRPRRSP 41
Db 45 REHAARATWRGDVHAR-GSRGCGGGGTRSLMAHGGPRARA 86

RESULT 9
P79522 PRELIMINARY; PRT; 253 AA.
ID P79522

```

```
AC P79522;
DT 01-MAY-1997 (TReMBLrel. 03, Created)
DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE MHC class I region proline rich protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wei H., Weissman S.M.;
RT "Human Proline Rich Sequence (CAT 56) from the MHC Class I Region.";
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U63336; AAB40147.1; -.
DR InterPro; IPR000571; Znf.CCCH.
DR Pfam; PF00642; zif-CCCH; 1.
DR SMART; SM00356; Znf.C3H1; 1.
DR SEQUENCE 253 AA; 27826 MW; C23696FAA0C36AC5 CRC64;
SQ SEQUENCE 253 AA; 27826 MW; C23696FAA0C36AC5 CRC64;

Query Match 22.5%; Score 56; DB 7; Length 253;
Best Local Similarity 36.7%; Pred. No. 18;
Matches 11; Conservative 2; Mismatches 17; Indels 0; Gaps 0;

QY 13 DLSWRLDAEGCGTPPSNRRAVKRHKPRSP 42
DB 25 EVEWREPHQCCPPTPRKRRKNRVLPHYP 54

RESULT 10
Q9MA52 PRELIMINARY; PRT; 463 AA.
AC Q9MA52;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE F22F7.17 protein.
GN F22F7.17.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B., Wu D.,
RA Maiti R., Ronning C.M., Koo H., Fujii C.Y., Utterback T.R.,
RA Barnstead M.E., Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome III BAC F22F7 genomic sequence.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC009606; AAF64543.1; -.
DR SEQUENCE 463 AA; 54260 MW; 6F104DD00D99E8BA CRC64;

Query Match 22.5%; Score 56; DB 10; Length 463;
Best Local Similarity 33.3%; Pred. No. 33;
Matches 14; Conservative 8; Mismatches 12; Indels 8; Gaps 2;

QY 11 CDD---LSWRLDAEGCGTPPSNR-----AVKHKGRPRSPAL 44
DB 203 CDDWNLAQKMLNGCDPLRRRCILTRASMTYQKPYPIESL 244 ,

RESULT 11
Q9BVL4 PRELIMINARY; PRT; 518 AA.
AC Q9BVL4;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Hypothetical 58.0 kDa protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Strausberg R.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC001099; AAH01099.1; -.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR InterPro; IPR003846; UPF0061.
DR Pfam; PF02696; UPF0061; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 518 AA; 58004 MW; B8466C79D9CDB3C1 CRC64;

Query Match 22.5%; Score 56; DB 4; Length 518;
Best Local Similarity 47.6%; Pred. No. 37;
Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 16 WRLDAEGCGTPPSNRRAVKRHK 36
DB 23 WELQLKGAGTPPFSROADGRK 43

RESULT 12
O83055 PRELIMINARY; PRT; 660 AA.
AC O83055;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE TPR protein B (TPRB).
GN TP0011.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
spirochete.";
RL Science 281:375-388(1998).
DR EMBL; AE001186; AAC65006.1; -.
DR TIGR; TP0011; -.
DR InterPro; IPR003872; MOSP_C.
DR InterPro; IPR003857; MOSP_Nterm.
DR Pfam; PF02722; MOSP_C; 1.
DR Pfam; PF02707; MOSP_N; 1.
KW Complete proteome.
SQ SEQUENCE 660 AA; 71106 MW; 11C5E0907ADE8A82 CRC64;

Query Match 22.5%; Score 56; DB 16; Length 660;
Best Local Similarity 37.0%; Pred. No. 48;
Matches 10; Conservative 7; Mismatches 8; Indels 2; Gaps 1;

QY 10 GCDDLSWRL--DAEGCGTPPSNRVAKH 34
DB 40 GSAELSGVYFDAEGASPVTAGKSINH 66

RESULT 13
Q9DBC0 PRELIMINARY; PRT; 664 AA.
ID Q9DBC0
AC Q9DBC0;
```

```
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)
DE 1300018J18RIK protein.
GN 1300018J18RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kaniya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK005048; BAB23774.1;
DR MGD: MGI:1919007; 1300018J18RIK.
DR InterPro: IPR002086; Aldehyde_dehydr.
DR InterPro: IPR003846; UPF0061.
DR Pfam: PF02696; UPF0061; 1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
SQ SEQUENCE 664 AA; 73853 MW; F27B8D90262A8AE2 CRC64;

Query Match 22.5%; Score 56; DB 11; Length 664;
Best Local Similarity 47.6%; Pred. No. 48;
Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 16 WRLDAEGCGPTPSNRVAKHRK 36
| | : | | | | : | |
DB 172 WELQLKGAGTPTFSRQADGRK 192

RESULT 14
Q9UGX5 PRELIMINARY; PRT; 770 AA.
AC Q9UGX5;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE DJ402G11.5 (novel protein similar to yeast and bacterial predicted
DE proteins).
DE proteins).
GN DJ402G11.5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Coville G.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL022328; CAB63045.1;
DR InterPro: IPR002086; Aldehyde_dehydr.
DR InterPro: IPR003846; UPF0061.
DR Pfam: PF02696; UPF0061; 1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
```

```
SQ SEQUENCE 770 AA; 83424 MW; 76491249F0C65840 CRC64;

Query Match 22.5%; Score 56; DB 4; Length 770;
Best Local Similarity 47.6%; Pred. No. 56;
Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 16 WRLDAEGCGPTPSNRVAKHRK 36
| | : | | | | : | |
DB 171 WELQLKGAGTPTFSRQADGRK 191

RESULT 15
Q95U05 PRELIMINARY; PRT; 922 AA.
AC Q95U05;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE GH168589p.
GN CG7699.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuno J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY058401; AAL13630.1;
DR Flybase: FBgn0033560; CG7699.
SQ SEQUENCE 922 AA; 99729 MW; 12ED56F6834C7CEC CRC64;

Query Match 22.5%; Score 56; DB 5; Length 922;
Best Local Similarity 40.5%; Pred. No. 68;
Matches 15; Conservative 6; Mismatches 6; Indels 10; Gaps 2;

QY 12 DDLRWLDAEGCGPTPSNRNRA----VKHRKPRPRSPAL 44
| | | | | : | | | | | : | | | | |
DB 504 DDLFSRRSSE-----ENLAGIEIVEHPQPKPLSPAI 534

Search completed: April 2, 2003, 15:53:51
Job time : 7.0137 secs
```

**THIS PAGE BLANK (USPTO)**